RISK AND TRADEOFFS

Lara Buchak, UC Berkeley

1. Introduction

Decision theories are theories of instrumental rationality: they formalize constraints of consistency between rational agents' ends and the means they take to arrive at these ends. We can model the possible actions an agent might take each as a gamble whose outcomes depend on the state of the world: for example, the action of not bringing an umbrella is a gamble that results in getting wet if it rains and staying dry if it doesn't. Decision theory places constraints on the structure of rational agents' preferences among the actions available to them and, as a result, can represent the beliefs and desires of any agent who meets these constraints by precise numerical values.

The prevailing view is that *subjective expected utility theory*, which dictates that agents prefer gambles with the highest expected utility, is the correct theory of practical rationality. That is, subjective expected utility theory (hereafter, EU theory) is thought to characterize the preferences of all rational decision makers. And yet, there are some preferences that violate EU theory that seem both intuitively appealing and prima facie consistent. An important group of these preferences stem from how ordinary decision makers take risk into account: ordinary decision makers seem to care about "global" properties of gambles, but EU theory rules out their doing so.

If one is sympathetic to the general aim of decision theory, there are three potential lines of response to the fact that EU theory does not capture the way that many people take risk into account when forming preferences among gambles. The first is to claim that contrary to initial appearances, expected utility theory can represent agents who care about global properties, by re-describing the outcomes that agents face. The second response is to claim that while many people care about global properties (and that these patterns of preferences cannot be represented by the theory), these people are simply not rational in doing so. I think that neither of these responses can succeed. I advocate a third response: modifying our normative theory to broaden the range of rationally permissible preferences. In particular, I advocate broadening the set of attitudes towards risk that count as rationally permissible. Although I won't directly argue against the first two responses, formulating an alternative will be important to evaluating them, since we need to know what it is that agents are doing when they systematically violate EU theory. In this paper, I will explain my alternative theory, and I will in particular explain how it does a better job at explicating the components of instrumental rationality than does EU theory.

2. Risk, EU Theory, and Practical Rationality

I begin by briefly explaining subjective expected utility theory; explaining how it must analyze the phenomenon of risk aversion; and showing that as a result, EU theory cannot capture certain

preferences that many people have. I will then argue that this problem arises for EU theory because it neglects an important component of practical rationality.

EU theory says that rational agents maximize expected utility: they prefer the act with the highest mathematical expectation of utility, relative to their utility and credence functions. So if we think of an act as a gamble that yields a particular outcome in a particular state of the world—for example, g = {O_1, E_1; O_2, E_2; … ; O_n, E_n} is the act that yields O_i if E_i is true, for each i —then the value of this act is:

$$\text{EU(g)} = \sum_{i=1}^{n} p(E_i)u(O_i)$$

If an agent (weakly) prefers f to g, then EU(f) > (≥) EU(g). So utility and credence are linked to rational preferences in the following way: if we know what an agent's utility function and credence function are, we can say what she ought to prefer. They are also linked in another way that will be of central interest in this paper: if we know an agent's preferences, and if these preferences conform to the axioms of EU theory, then we can determine her credence function uniquely and her utility function up to positive affine transformation:[1] we can represent her as an expected utility maximizer relative to a some particular p and u. It is crucial for the EU theorist that the preferences of all rational agents can be represented in this way.

It is uncontroversial that many people's preferences display risk aversion in the following sense: an individual would rather have $50 than a fair coin flip between $0 and $100, and, in general, would prefer to receive $$z$ rather than to receive a gamble between $$x$ and $$y$ that will yield $$z$ on average.[2] If the agent is representable as an EU maximizer, then it must be that u($50) – u($0) > u($100) – u($50), i.e., that getting the first $50 makes more of a utility difference than getting the second $50 does. More generally, her utility function in money must diminish marginally:

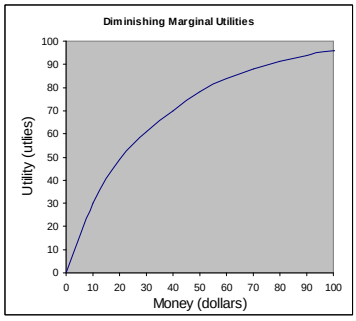

Diagram 1: Diminishing Marginal Utility

And vice versa: when a utility function is concave, a risky gamble (say, the fair coin-flip between $0 and $100) will always have an expected utility that is less than the utility of its expected dollar value ($50).[3] On EU theory, then, aversion to risk is equivalent to diminishing marginal utility.

Intuitively, though, there are two different psychological phenomena that could give rise to risk-averse behavior. On the one hand, how much one values additional amounts of money might diminish the more money one has. As an extreme case of this, consider Alice, who needs exactly $50 for a bus ticket, and doesn't have much to buy beyond that. On the other hand, one might value small amounts of money linearly, but care about other properties of the gamble besides its average utility value: for example, the minimum value it might yield, the maximum, or the spread of possible outcomes. In other words, one might be sensitive to *global properties*. Consider Bob, who gets as much pleasure out of the first $50 as the second, but would rather guarantee himself $50 than risk having nothing for the possibility of $100. Both Alice and Bob prefer $50 to the coin-flip, and the EU theorist must interpret both agents as having a non-linear utility function, on the basis of this preference.

What is the relationship between an agent's psychology the utility function that is derived from her preferences? There are two views about this: on the *realistic* picture, the utility function corresponds to something "in the head": it is a measure of how much an agent desires or prefers something or the degree of satisfaction that having it could bring her; that is, it represents some pre-existing value that the agent has. So for the realist, EU theory will have misinterpreted Bob, since Bob's strength of desire for money is linear.

On the *constructivist* picture, which seems to have more widespread endorsement among contemporary philosophers,[4] the utility function is not meant to measure strength of preference, or goodness, or any quantity that exists independently in the head or in the world. The constructivist thinks that we cannot give any independent content to these notions. Utility is instead a theoretical construct from preferences: it is the quantity whose mathematical expectation an agent maximizes. So the constructivist won't necessarily care about the agent's reasons for the preferences she has. However, these reasons will turn out to matter, because if the formalist EU theorist misses an important fact about the agent's psychology, then although he will be able to explain an isolated preference (like that for $50 over the $0/$100 coin-flip), his explanation will commit the agent to having preferences that she does not in fact have and will therefore fail to represent her.

Matthew Rabin presents a "calibration theorem" to show that in order to describe the preferences of decision makers that display risk aversion in modest-stakes gambles, EU theory is committed to absurd conclusions about preferences between gambles when the stakes are higher (absurd in the sense that no one actually has these preferneces). As mentioned, EU theory must interpret modest stakes risk-aversion as entailing a concave utility function. Rabin's results assume nothing about the utility function except that it continues to be concave in higher stakes, and so doesn't, for example, have an inflection point above which it increases marginally. Here are some examples of the sorts of conclusions EU theory is

committed to.[5] If an agent prefers not to take a fair coin-flip between losing $100 and gaining $110 (that is, if she prefers a sure-thing $0 to the coin-flip), regardless of her initial wealth level, then she must also prefer not to take a coin-flip between losing $1,000 and gaining *any amount* of money. Similarly, if an agent prefers not to take a coin-flip between losing $1,000 and gaining $1050 for any initial wealth level, then she will also prefer not to take a coin-flip between losing $20,000 and gaining any amount of money. Furthermore, if an agent prefers not to take a coin-flip between losing $100 and gaining $105 as long as her lifetime wealth is less than $350,000, then from an initial wealth level of $340,000, she will turn down a coin-flip between losing $4,000 and gaining $635,670. In other words, she will prefer a sure-thing $340,000 to the gamble {$339,600, 0.5; $975,670, 0.5}.[6]

Rabin's results are problematic for both the realist and constructivist EU theorist: if most people have the modest-stakes preferences but lack the high-stakes preferences that 'follow' from them, then EU theory with a diminishing marginal utility function will fail to represent most people. In case the reader is worried that Rabin's results rely on knowing a lot of the agent's preferences, there are also examples of preferences that EU theory (under either interpretation) cannot account for that involve very few preferences. One example is Allais's famous paradox. Consider Maurice, who is presented with two hypothetical choices, each between two gambles.[7] He is first asked whether he would rather have L_1 or L_2:

L_1: $5,000,000 with probability 0.1, $0 otherwise.

L_2: $1,000,000 with probability 0.11, $0 otherwise.

He reasons that the minimum he stands to walk away with is the same either way, and there's not much difference in his chances of winning *some* money. So, since L_1 yields much higher winnings at only slightly lower odds, he decides he would rather have L_1. He is then asked whether he would rather have L_3 or L_4:

L_3: $1,000,000 with probability 0.89, $5,000,000 with probability 0.1, $0 otherwise.

L_4: $1,000,000 with probability 1.

He reasons that the minimum amount that he stands to win in L_4 is a great deal higher than the minimum amount he stands to win in L_3, and that although L_3 comes with the possibility of much higher winnings, this fact is not enough to offset the possibility of choosing L_3 and ending up with nothing. So he decides he would rather have L_4. Most people, like Maurice, prefer L_1 to L_2 and L_4 to L_3. However, there is no way to assign utility values to $0, $1m, and $5m such that L_1 has a higher expected utility than L_2 and L_4 has a higher expected utility than L_3 and therefore these preferences cannot be represented as maximizing expected utility.[8] Allais's example does not require any assumptions about an agent's psychology – it relies only on the agent having the two preferences mentioned – and so again presents a problem for both the realist and the constructivist EU theorist.

Most people have preferences like those that Allais and Rabin show cannot be captured by EU theory; and there are many other examples of preferences that EU theory cannot capture. The reason EU theory fails to capture the preferences in the Rabin and Allais examples is that it fails to separate two different sorts of reasons for risk averse preferences: local considerations about outcomes, like those that Alice advanced in order to determine that she prefers $50 ("this particular amount of money is more valuable…") and global considerations about gambles as a whole, like those that Bob advanced in order to determine that he prefers $50 ("I would rather be guaranteed $50 than risk getting less for the possibility of getting more").

Why would an agent find this second kind of consideration relevant to decision making? Let us examine the idea that decision theory formalizes and precisifies means-ends rationality. We are presented with an agent who wants some particular end and can achieve that end through a particular means. Or, more precisely, with an agent who is faced with a choice among means that lead to different ends, which he values to different degrees. To figure out what to do, the agent must make a judgment about which ends he cares about, and how much: this is what the utility function captures.[9] In typical cases, none of the means available to the agent will lead with certainty some particular end, so he must also make a judgment about the likely result of each of his possible actions. This judgment is captured by the subjective probability function. Expected utility theory makes precise these two components of means-ends reasoning: how much one values various ends, and which courses of action are likely to realize these ends.

But this can't be the whole story: what we've said so far is not enough for an agent to reason to a unique decision, and so we can't have captured all that is relevant to decision making. An agent might be faced with a choice between one action that guarantees that he will get something he desires somewhat and another action that might lead to something he strongly desires, but which is by no means guaranteed to do so. **Knowing how much he values the various ends involved and how likely each act is to lead to each end is not enough to determine what the agent should do in these cases: the agent must make a judgment not only about how much he cares about *particular* ends, and how effective his actions will be in realizing *each* of these ends, but about which sort of *strategy* to take towards realizing his ends *as a whole*: how to structure the realization of his aims.** This involves deciding whether to prioritize definitely ending up with something of some value or instead to prioritize possibly ending up with something of extraordinarily high value, and by how much: specifically, he must decide the extent to which he is generally willing to accept the risk of something worse in exchange for the possibility of something better. This judgment corresponds to considering global or structural properties of gambles.

How should an agent trade off the fact that one means will realize some end for sure against the fact that another means has some small possibility of realizing some different end that he cares about more? This question won't be answered by consulting the probabilities of states or the utilities of ends. Two agents could attach the very same values to certain ends (various sums of money, say), and they could have the same beliefs about how likely various means are to achieve their ends. And yet, one agent might think his preferred strategy for generally fulfilling his desires involves taking a gamble that has a small chance of a very high payoff, whereas the other might think that he can more effectively achieve *this same general goal* by taking a gamble with a high chance of a moderate payoff. Knowing they can only achieve some of their aims, these agents have two different ways to structure the realization of (some of) these aims.

This dimension of instrumental reasoning is the dimension of evaluation that standard decision theory has ignored. To be precise, it hasn't ignored it but rather supposed that there is a single correct answer for all rational agents: one ought to take actions that have higher utility on average, regardless of the spread of possibilities. There may or may not be good arguments for this, but we are not in a position to address them before we get clear on what exactly agents are doing when they answer the question differently, and how this relates to practical reasoning. The aim of this paper is to make this clear.

3. An Alternative Theory

To explain the alternative theory of instrumental rationality I endorse, I will start with the case of gambles with only two outcomes: gambles of the form $\{O_1$ if E, O_2 if ~E$\}$. As mentioned, the EU of such a gamble is $p(E)u(O_1) + p(\sim E)u(O_2)$. We can state this equivalently as $u(O_2) + p(E)[u(O_1) - u(O_2)]$. Taking O_2 to be the less (or equally) desirable outcome, this says that EU is calculated by taking the minimum utility value the gamble might yield, and adding to it the potential gain above the minimum (the difference between the high value and the low value), weighted by the probability of receiving that gain. For example, the value of the \$0/\$100 coin-flip will be $u(\$0) + (0.5)[u(\$100) - u(\$0)]$.

Again, this implies that while it is up to agents themselves how valuable each outcome is and how likely they believe each state is to obtain, these two evaluations are of set significance to the overall value of a gamble. If two decision makers agree on the values of various outcomes and on the probabilities involved, they must evaluate gambles in exactly the same way: their preference ordering must be exactly the same. However, it is plausible to think that some people might be more cautious (or prudent) than others, again, for purely instrumental reasons: because they think that guaranteeing themselves something of moderate value is a better strategy for satisfying their general aims of getting some of the things that they value than is making a very high amount (merely) possible. More realistically, the minimum value won't always trump the maximum in their considerations, but it will weigh more heavily. Or,

alternatively, an agent might be incautious: the possibility of the gamble yielding the maximum will weigh more heavily in the estimation of its value than its minimum will, even if these two outcomes are equally likely. So, it is plausible that two agents who attach the same value as each other to $100 and $0 will not both attach the same value to the coin-flip. The cautious agent will take the fact that he has a 50% chance of winning the better prize to be a weaker consideration than it is for the incautious agent. Thus, in addition to having different attitudes towards outcomes, and different evaluations of likelihoods, two agents might have different attitudes towards ways of potentially attaining some of these outcomes.

In EU theory, the potential gain above the minimum is weighted by the probability of realizing that gain. But this is too restrictive: a potential gain over the minimum *might* improve a gamble by the size of the gain multiplied by the probability of receiving that gain, but it might instead improve it by more or by less, depending on the agent's strategy for realizing his aims. Of course, the probability and the size of the improvement will be relevant: some particular probability of a larger gain rather than a smaller gain will be better, and the higher the probability of some particular gain, the better. Therefore, I propose that the possibility of a potential gain over the minimum improves the gamble above its minimum value by the size of the gain multiplied by a *function of* the probability of realizing that gain, instead of by its bare probability. This function measures the agent's attitude towards risk in the "global properties" sense: it measures how an agent takes into account the possibility of doing better than the worst-case scenario. So the value of a two-outcome gamble will be its low value plus the interval between the low value and the high value, weighted by a function of the probability of getting the high value. Put formally, we might calculate the *risk-weighted expected utility (REU)* of a gamble $\{O_2$ if E, O_1 if ~E$\}$, where $u(O_1) \leq u(O_2)$, to be $u(O_1) + \mathbf{r(}p(E)\mathbf{)}[u(O_2) - u(O_1)]$, where r is the agent's "risk function" or "weighting function," adhering to the constraints $r(0) = 0$, $r(1) = 1$, r is non-decreasing, and $0 \leq r(p) \leq 1$ for all p.

Note that this is equivalent to $r(p(E))u(O_2) + (1 - r(p(E))u(O_1)$. So if the function has a high value for some p, then the value of the better outcome will count for a lot in the agent's evaluation of the gamble, and if it has a low value for some p, then the value of the worse outcome will count for a lot. This formulation also makes it clear how an agent's evaluation of gambles rests on factors that are irreducibly global: the amount by which each outcome gets weighted will depend on which outcome is the minimum.[10]

For example, for an agent who values money linearly and has a risk function of $r(p) = p^2$, the coin-flip will be worth $25: $u(\{\$0, \text{HEADS}; \$100, \text{TAILS}\}) = u(\$0) + (0.5)^2[u(\$100) - u(\$0)] = u(\$25)$.

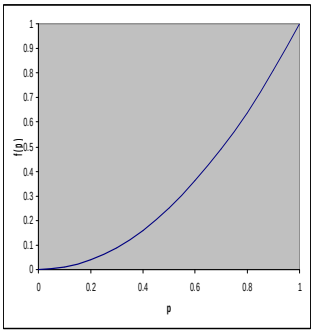

Diagram 2: Sample Risk Function: $r(p) = p^2$

And here we come to the crux of the difference between the psychological state that the standard theory thinks is properly called risk aversion and the psychological state that I think merits the term. *On EU theory, to be risk averse is to have a concave utility function. On a theory like mine, to be risk averse is to have a convex risk function.*[11] The intuition behind the diminishing marginal utility analysis of risk aversion was that adding money to an outcome is of less value the more money it already contains; or that getting an additional good is of less value if one already has some other good. The intuition behind my analysis of risk aversion is that adding *probability* to an outcome is of more value the more likely that outcome already is to obtain. Risk averters prefer to "get to certainty," so to speak. Of course, theories like mine allow that the utility function is concave (or, indeed, any shape). But I claim that this feature, which describes how an agent evaluates outcomes, pulls apart from his attitude towards risk properly called. So I claim that what we might appropriately describe as an agent's attitude towards risk is captured by the shape of his risk function.

There is a natural way to extend this theory to gambles with more than two possible outcomes. The way I've set up the risk-weighted expected utility equation emphasizes that an agent considers his possible gain above the minimum (the interval between the low outcome and the high outcome), and weights that gain by a factor which is a function of the probability of obtaining it, a function that depends on how he regards risk. Now consider a situation in which a gamble might result in one of *more than two* possible outcomes. It seems natural that he should consider the possible gain between each neighboring pair of outcomes and his chance of arriving at the higher outcome or better, and, again, subjectively determine how much that chance of getting something better adds to the value of the gamble.

One way to state the value of a gamble with more than two outcomes for a standard EU maximizer is as follows. Start with the minimum value. Next, add the interval difference between this value and the next highest value, weighted by the probability of getting at least that higher value. Then add the interval difference between this value and the next highest value, weighted by the probability of getting at least *that* value. And so forth. Just as we replaced subjective probabilities by subjective weights of subjective probabilities in the two-outcome case, we can do so in this case. So the value of a

gamble for the REU maximizer will be determined by following this same procedure but instead weighting by a function of the probability at each juncture.

For example, consider the gamble that yields \$1 with probability ½, \$2 with probability ¼, and \$4 with probability ¼. The agent will get at least \$1 for certain, and he has a ½ probability of making at least \$1 more. Furthermore, he has a ¼ probability of making at least \$2 beyond *that*. So the REU of the gamble will be u(\$1) + r(½)[u(\$2) – u(\$1)] + r(¼)[u(\$4) – u(\$2)].

So the gamble g = {O_1 if E_1; O_2 if E_2; … ; O_n if E_n}, where $u(O_1) \leq \ldots \leq u(O_n)$, is valued under expected utility theory as as $\sum_{i=i}^{n} p(E_i)u(O_i)$, which is equivalent to:

EU(g) =

$$u(O_1) + (\sum_{i=2}^{n} p(E_i))(u(O_2) - u(O_1)) + (\sum_{i=3}^{n} p(E_i))(u(O_3) - u(O_2)) + \ldots + p(E_n)(u(O_n) - u(O_{n-1}))$$

And that same gamble will be valued under risk-weighted expected utility theory as follows:

REU(g) =

$$u(O_1) + r(\sum_{i=2}^{n} p(E_i))(u(O_2) - u(O_1)) + r(\sum_{i=3}^{n} p(E_i))(u(O_3) - u(O_2)) + \ldots + r(p(E_n))(u(O_n) - u(O_{n-1}))$$

We can now see how the standard Allais preferences are captured by REU theory: they maximize risk-weighted expected utility only if r is convex.[12]

This functional form mirrors the "rank-dependent" approach in non-expected utility theories discovered by economists around the 1980s, in which the agent maximizes a sum of utility values of outcomes, weighted by a factor that is related to the probability of that outcome but that depends on the outcome's rank among possible outcomes. In particular, two of these theories are formally equivalent to REU theory when we abstract away from what their "weighting factor" is a function of. The first is Choquet expected utility (CEU), due to David Schmeidler and Itzhak Gilboa,[13] and the second is anticipated utility (AU), due to John Quiggin.[14] However, CEU employs a weighting function of states, not of probabilities of states: it does not include an agent's judgments about probabilities at all. Indeed, it is meant to apply to decision making under uncertainty, in which agents do not always have sharp probability judgments.[15] AU does attach decision weights to probabilities, but it uses an "objective" probability function: it takes the probabilities as given. My formulation allows that an agent attaches subjective probabilities to states and then employs a weighting function of these probabilities. This is crucial for philosophers working in decision theory, since philosophers are particularly interested extracting beliefs (as well as desires) from preferences.[16]

If we set r(p) = p, we get the standard subjective expected utility equation. And again, for agents who are cautious – agents with convex r-functions – the possibility of getting higher than the minimum will improve the value of the gamble less than it will for the expected utility maximizer. The most extreme case of this is the maximinimizer, who simply takes the gamble with the highest minimum. He can be represented using r(p) = {0 if $p \neq 1$, 1 if $p = 1$}. And for agents who are incautious – agents with concave r-functions – the possibility of getting higher than the minimum will improve the value of the gamble more. The maximaximizer, who takes the gamble with the highest maximum, can be represented using r(p) = {0 if $p = 0$, 1 if $p \neq 0$}. The REU equation also ensures that the value of a gamble is always at least its minimum and at most its maximum, and, since r is non-decreasing, that improving the probability of getting a good outcome will never make a gamble worse (preferences respect weak stochastic dominance).

What ingredient of instrumental rationality does the risk function represent? The utility function is traditionally supposed to represent desire, and the probability function belief – both familiar propositional attitudes. We try to make beliefs "fit the world," and we try to make the world fit our desires. But the risk function is neither of these things: it does not quantify how we see the world – it does not, for example, measure the strength of an agent's belief that things will go well or poorly for him – and it does not describe how we would like the world to be. It is not a belief about how much risk one should tolerate, nor is it a desire for more or less risk. The risk function corresponds to neither beliefs nor desires. Instead, it measures how an agent structures the realization of his aims. (We will see in the next section exactly how it does this.)

Thus the agent subjectively determines *three* things: which ends he wants, how likely various actions are to lead to various ends, and the extent to which he is generally willing to accept the risk of something worse in exchange for the possibility of something better. First, like the standard theory, REU theory allows agents to attach subjective values to outcomes. It is up to agents themselves to choose their ends, and hence, my theory includes a subjective utility function, which is not necessarily linear in money. Second, also like the standard theory, it allows them to assess the probability of some particular act leading to some particular result, and hence, my theory includes a subjective probability function. Third, *unlike* the standard theory, it allows them to subjectively judge which sorts of means are more effective at fulfilling their ends as a whole. It allows them to judge which gamble better realizes their aim of getting more money (which includes their particular ends of, say, getting $50 or, perhaps better by twice, getting $100). It is up to them whether they will better fulfill their goals by guaranteeing themselves a high minimum or by allowing themselves the possibility of some high maximum. And it is up to them how these two features of gambles trade off, e.g., how much possibly doing better than the

minimum is worth. Hence, my theory includes a subjective risk function, which is not necessarily linear in probability.

To put this point another way: every agent has beliefs, desires, and a norm for translating these two things into preferences. EU theory assumes r(p) = p is the correct norm. But I claim that just as there may be no uniquely correct utility function (in the spirit of Hume, as long as one is consistent, one can have any preferences one wants), I claim that there is also no uniquely correct norm for rational agents.

4. From Preferences to Beliefs and Desires

I have so far been focusing on the question of how an agent might aggregate her beliefs (credence function) and desires (utility function) to arrive at a single value for an act. I've claimed that agents need not aggregate according to the expected utility principle, but instead might weight the values of outcomes by a function of their probabilities. Thus we might model decision makers as using a more general decision rule, which includes a utility function, a credence function, and a risk function. However, the question that has received the most attention in philosophy is not how we might arrive at preferences once we know an agent's beliefs and desires, but rather how we might extract an agent's beliefs and desires from her preferences. Specifically, decision theorists have been interested in what restrictions on preferences will allow us to fix unique beliefs and desires. So the question that arises for my theory is what restrictions are needed in order to extract beliefs, desires, and attitudes towards risk.

As mentioned, a utility and probability function represent an agent under EU theory just in case for all acts f and g, the agent prefers f to g iff $EU(f) \leq EU(g)$, where expected utility is calculated relative to the agent's subjective probability function of states. A *representation theorem* spells out a set of axioms such that if an agent's preferences obey these axioms, then she will be representable under EU theory by a unique probability function and a utility function that is unique up to positive affine transformation: she will be an EU maximizer relative to these functions.[17]

Representation theorems are important in decision theory, but their upshot depends on the use to which decision theory is put. There are at least two very different ways in which decision theory has been used, which I refer to as the *prescriptive* use and the *interpretive* use. When the theory is taken prescriptively, an agent uses it to identify the choice he should make or the preferences he should have; or the decision theorist uses the theory to assess whether the agent's choices and preferences are rational.

The agent himself can use decision theory prescriptively in at least two ways. First, if an agent has already formed preferences over enough items, he can look to decision theory to tell him the preferences he should have over other items; that is, to tell him how to attain his ends of getting, on balance, things that he (already) more strongly prefers. Second, if an agent realizes that his preferences are not in accord with decision theory, then he can conclude that he has done something wrong and,

insofar as he is rational, that he should alter his preferences so that they do so accord. In addition, the *theorist* using decision theory prescriptively ascertains whether the agent's choices in fact accord with the theory, and it is by this criterion that she judges whether the agent's preferences are rational.

Representation theorems state the conditions under which an agent can count as an EU maximizer, and thus the conditions under which an agent's preferences are rational (on the standard theory). Therefore, they are useful for prescriptive decision theory because they provide a criterion for determining when an agent has irrational preferences that doesn't require knowing his precise numerical values. Furthermore, this criterion can be employed if we think there are no precise numerical values to know (aside from those that result from the theorem), so they are especially useful to the constructivist. For the constructivist, rationality just is conformity to the axioms of decision theory, and it is a convenience that this also guarantees representability as an expected utility maximizer. Thus, representation theorems are useful because they allow us to refocus the debate about rationality: instead of arguing that a rational agent ought to maximize expected utility because, say, he ought to care only about average value, the EU theorist can argue that a rational agent ought to conform to the axioms.

In contrast to prescriptive decision theory, a portion of the modern philosophical literature treats decision theory *interpetively*: not as a useful guide to an agent's own decisions, but rather as a framework to interpret an agent's desires, his beliefs, and perhaps even the options that he takes himself to be deciding among. The interpretive use of decision theory arises in response to a worry about how to discover what an agent believes and desires, given that we have no direct access to these mental states – and, if constructivism (or a version of realism on which one's desires are opaque) is true, neither do the agents themselves, since these states cannot be discovered by introspection. However, it seems that it is relatively easy to discover agents' preferences: preferences do manifest themselves directly (if not perfectly) in behavior, and are ordinarily open to introspection.

It should be clear how representation theorems are useful to interpretive theorists. If an agent's preferences obey the axioms of EU theory, then the interpretive theorist can start with the agent's (observable) preferences and derive the probability function and utility function that represent her. It should also be clear why it is important that the theorems result in *unique* probability and utility functions. If there were multiple <p, u> pairs that each could represent the agent as an expected utility maximizer, we wouldn't know which way of representing the agent accurately captures "her" beliefs and desires.[18]

We can see that representation theorems are crucial to decision theory, so any alternative to EU theory needs a representation theorem if it can hope to serve the purposes EU theory is traditionally put to. Furthermore, comparing the axioms of an EU representation theorem to those of an alternative will allow us to see the difference between what each theory requires of rational agents. In the remainder of

this paper, I will present a representation theorem for my theory (though I will not include the proof here).[19] In particular, I will present a set of axioms such that if a decision maker's preferences obey these axioms, we will be able to determine a unique probability function, a unique risk function, and a utility function that is unique up to positive affine transformation such that an agent maximizes REU relative to these three functions. My aim here is to show how REU theory captures what it is to be risk-averse at the level of preferences, by contrasting the axioms of my theorem and the (stronger) axioms of the analogous representation theorem for expected utility theory. This will provide a way to frame the debate about whether REU maximizers are rational around the question of whether agents ought to obey the axioms of EU theory or only the weaker axioms of my theory.

5. Representation Theorem

My theorem draws on two other results, one by Veronika Kobberling and Peter Wakker (hereafter KW) and by Mark Machina and David Schmeidler (hereafter MS).[20] Kobberling and Wakker prove a representation theorem for another "rank-dependent theory," CEU theory. CEU, like my theory, applies to "Savage" acts, in which outcomes are tied to events whose probabilities are not given. However, as mentioned, CEU does not represent the agent as having a function that assigns *probabilities* to events, and thus the representation theorem for CEU does not provide us with a way of extracting the agent's degrees of belief from his preferences. Machina and Schmeidler give conditions under which an agent can be represented as probabilistically sophisticated – as having a unique probability function relative to which his preferences respect stochastic dominance – and as maximizing *some* value function, but does not allow us to determine the values of outcomes aside from the gambles they are embedded in. Combining their results, I give conditions under which we can represent an agent as a probabilistically sophisticated decision maker maximizing the specific function that this paper is concerned with: that is, giving conditions under which we can extract from an agent's preferences a probability function, a utility function, and a *function that represents how he structures the realization of his aims in the face of risk.* Thus the set of axioms I will use in my representation theorem are a combination of Kobberling and Wakker's and Machina and Schmeidler's axioms, strictly stronger than either set of axioms.

I start by explaining the spaces and relations we are dealing with.[21] The **state space** is a set of states S = {…, s, …}, whose subsets are called events. The **event space**, EE, is the set of all subsets of S. Since I want to represent agents who have preferences over not just monetary outcomes but discrete goods and, indeed, over fully specified states of the world, it is important that the outcome space be general. Therefore, the outcome space is merely a set of arbitrary outcomes X = {…, x, …}. I follow Savage (1954) in defining the entities an agent has preferences over as "acts" that yield a known outcome in each state. The **act space** A = {…, f(.), g(.), …} is thus the set of all finite-valued functions from S to

X, where the inverse of each outcome $f^{-1}(x)$ is the set of states that yields that outcome: $f^{-1}(x) \in EE$. So for any act $f \in A$, there is some partition of the event space EE into $\{E_1, \ldots E_n\}$ and some finite set of outcomes $Y \subseteq X$ such that f can be thought of as a member of Y^n. And as long as f(s) is the same for all $s \in E_i$, we can write $f(E_i)$ as shorthand for "f(s) such that $s \in E_i$."

For any fixed finite partition of events $M = \{E_1, \ldots, E_n\}$, all the acts on those events will form a subset $A_M \subseteq A$. Thus, A_M contains all the acts that yield, for each event in the partition, the same act for all states in that event: $A_M = \{f \in A \mid (\forall E_i \in M)(\exists x \in X)(\forall s \in E_i)(f(s) = x)\}$. An upshot is that for all acts in A_M, we can determine the outcome of the act by knowing which event in M obtains: we needn't know the state of the world in a more fine-grained way.

The **preference relation** $\geq$ is a two-place relation over the act space. This gives rise to the indifference relation and the strict preference relation: $f \sim g$ iff $f \geq g$ and $f \geq g$; and $f > g$ if $f \geq g$ and $\neg(g \geq f)$.

For all $x \in X$, fx denotes the constant function (f(i) = x for all i). I will sometimes use expressions that technically denote outcomes as the relata of < when to use "f" would be cumbersome, but these should always be read as denoting the constant function yielding that outcome in every state. Furthermore, $x_E f$ denotes the function that agrees with f on all states not contained in E, and yields x on any state contained in E. That is, $x_E f(s) = \{x \text{ if } s \in E; f(s) \text{ if } s \notin E\}$. Likewise, for disjoint E_1 and E_2 in EE, $x_{E1}y_{E2}f$ is the function that agrees with f on all states not contained in E_1 and E_2, and yields x on E_1 and y on E_2. We say that an event E is **null** on $F \subseteq A$ if the agent is indifferent between any pair of acts which differ only on E: $x_E f \sim f$ for all $x_E f, f \in F$.[22]

The concepts in this paragraph and the next are important in Kobberling and Wakker's result. Two acts f and g are **comonotonic** if there are no states $s_1, s_2 \in S$ such that $f(s_1) > f(s_2)$ and $g(s_1) < g(s_2)$. This is equivalent to the claim that there are no events $E_1, E_2 \in EE$ such that $f(E_1) > f(E_2)$ and $g(E_1) < g(E_2)$. The acts f and g order the states (and, consequently, the events) in the same way, so to speak: if s_1 leads to a strictly more preferred outcome than s_2 for act f, then s_1 does not lead to a strictly less preferred outcome than s_2 for act g. We say that a subset C of A_M is a **comoncone** if all the acts in C order the events in the same way: for example, the set of all acts on coin-flips in which the heads outcome is as good as or better than the tails outcome forms a comoncone. Formally, as Kobberling and Wakker define it, take any fixed partition of events $M = \{E_1, \ldots, E_n\}$. A permutation ρ from {1, …, n} to {1, …, n} is a *rank-ordering* permutation of f if $f(E_{\rho(1)}) \geq \ldots \geq f(E_{\rho(n)})$. So a comoncone is a subset C of A_M that is rank-ordered by a given permutation: $C = \{f \in A_M \mid f(E_{\rho(1)}) \geq \ldots \geq f(E_{\rho(n)})\}$ for some ρ. For each fixed partition of events of size n, there are n! comoncones.[23] This concept will become important in arguing that REU theory is a

better analysis of rational preference than EU theory, and will be discussed extensively in the next section.

We say that outcomes $x^1, x^2, \ldots$ form a **standard sequence** on $F \subseteq A$ if there exist an act $f \in F$, events $E_i \neq E_j$ that are non-null on F, and outcomes y, z with $\neg(y \sim z)$ such that $(x^{k+1})_{Ei}(y)_{Ej}f \sim (x^k)_{Ei}(z)_{Ej}f$, with all acts contained in F.[24] The idea behind a standard sequence is that the set of outcomes x^1, x^2, x^3, …, will be "equally spaced." (I should say: this is the interpretation we will be aiming for when we extract utility from preferences, but since we haven't stated the axioms yet the notion of a standard sequence doesn't yet have that meaning.) Since the agent is indifferent for each pair of gambles, and since each pair of gambles differs only in that the "left-hand" gamble offers y rather than z if E_j obtains, and offers x^{k+1} rather than x^k if E_i obtains, the latter tradeoff must exactly make up for the former. And since the possibility of x^{k+1} rather than x^k (if E_i) is enough to make up for y rather than z (if E_j) for each k, the difference between each x^{k+1} and x^k must be constant. Note that a standard sequence can be increasing or decreasing, and will be increasing if $z > y$ and decreasing if $y > z$. A standard sequence is bounded if there exist outcomes v and w such that $\forall i(v \geq x^i \geq w)$.

We are now in a position to define a relation that is important for Kobberling and Wakker's result and that also makes use of the idea that one tradeoff exactly makes up for another. For each partition M, we define the relation $\sim^*(F)$ for $F \subseteq A_M$ and outcomes $x, y, z, w \in X$ as follows:

$xy \sim^*(F) zw$

iff $\exists f,g \in F$ and $\exists E \in EE$ that is non-null on F such that $x_E f \sim y_E g$ and $z_E f \sim w_E g$,

where all four acts are contained in F.[25] Kobberling and Wakker explain the relation $\sim^*(F)$ as follows: "The interpretation is that receiving x instead of y apparently does the same as receiving z instead of w; i.e. it exactly offsets the receipt of the [f's] instead of the [g's] contingent on [¬E]."[26] The idea here is that if one gamble offers f if ~E obtains, whereas another gamble offers g if ~E obtains, then this is a point in favor of (let's say) the first gamble. So in order for an agent to be indifferent between the two gambles, there has to be some compensating point in favor of the second gamble: it has to offer a better outcome if E obtains. And it has to offer an outcome that is better by the right amount to exactly offset this point. Now let's assume that offering y rather than x (on E), and offering w rather than z (on E) both have this feature: they both exactly offset the fact that a gamble offers f rather than g (on ¬E). That is, if one gamble offers f on ~E, and a second gamble offers g on ~E, then this positive feature of the first gamble would be exactly offset if the first offered x on E and the second offered y on E – and it would be exactly offset if instead the first offered z on E and the second offered w on E. If this is the case, then there is some important relationship between x and y on the one hand and z and w on the other: there is a situation in which having the first member of each pair rather than the second both play **the same**

compensatory role. We might call this relationship ~* **tradeoff equality**. Following Kobberling and Wakker, I write xy ~*(C) zw if there exists a comonocone $F \subseteq A_M$ such that xy ~*(F) zw: that is, if x and y play the same compensatory role as z and w in some gambles in the same comoncone.

The relation ~*(F), and particularly ~*(C), will feature centrally in the representation theorem, because one important axiom will place restrictions on when it can hold, i.e., when two pairs of acts can play the same compensatory role. This relation will also play a crucial role in determining the (cardinal) value difference between outcomes from ordinal preferences. I will explain this more fully in the next section.

With the preliminaries out of the way, I can now present the axioms of REU theory, side-by-side with those of the analogous representation theorem for EU theory that Kobberling and Wakker spell out.[27]

EXPECTED UTILITY THEORY	RISK-WEIGHTED EXPECTED UTILITY THEORY
B1. Ordering: $\geq$ is complete, reflexive, and transitive.	**A1. Ordering** (MS P1): $\geq$ is complete, reflexive, and transitive.
B2. Nondegeneracy: There are at least two non-null states on A, and there exist outcomes x and y such that fx > fy.	**A2. Nondegeneracy** (MS P5): There exist outcomes x and y such that fx > fy.
B3. Weak (finite) monotonicity (KW 396): For any fixed partition of events $E_1, \ldots, E_n$ and acts $f(E_1, \ldots, E_n)$ on those events, if $f(E_j) \geq g(E_j)$ for all j, then $f \geq g$.	**A3. State-wise dominance:** If $f(s) \geq g(s)$ for all $s \in S$, then $f \geq g$. If $f(s) \geq g(s)$ for all $s \in S$ and $f(s) > g(s)$ for all $s \in E \subseteq S$, where E is non-null on A, then $f > g$.
B4. Solvability (KW 398): For any fixed partition of events $E_1, \ldots, E_n$, and for all acts $f(E_1, \ldots, E_n)$, $g(E_1, \ldots, E_n)$ on those events, outcomes x, y, and events E_i with $x_E f > g > y_E f$, there exists an "intermediate" outcome z such that $z_E f \sim g$.	**A4. Continuity** ((i)KW 398, Solvability, and (ii)MS P6 Small Event Continuity): (i) For any fixed partition of events $E_1, \ldots, E_n$, and for all acts $f(E_1, \ldots, E_n)$, $g(E_1, \ldots, E_n)$ on those events, outcomes x, y, and events E_i with $x_E f > g > y_E f$, there exists an "intermediate" outcome z such that $z_E f \sim g$. (ii) For all acts $f > g$ and outcome x, there exists a finite partition of events $\{E_1, \ldots, E_n\}$ such that for all i, $f > x_{Ei} g$ and $x_{Ei} f > g$
B5. Archimedean Axiom (KW 398): Every bounded standard sequence on A is finite.	**A5. Comonotonic Archimedean Axiom** (KW 398, 400): For each comoncone F, every bounded standard sequence on F is finite.
B6. Unrestricted Tradeoff Consistency (KW 397): Improving an outcome in any ~*(A) relationship breaks that relationship. In other words, xy ~*(A) zw and y' > y entails ¬(xy' ~*(A) zw).	**A6. Comonotonic Tradeoff** Consistency (KW 397, 400): Improving an outcome in any ~*(C) relationship breaks that relationship. In other words, xy ~*(C) zw and y' > y entails ¬(xy' ~*(C) zw).
	A7. Strong Comparative Probability (MS P4*): For all pairs of disjoint events E_1 and E_2, outcomes x' > x and y' > y, and acts $g, h \in A$, $x'_{E1}x_{E2}g \geq x_{E1}x'_{E2}g \Rightarrow y'_{E1}y_{E2}h \geq y_{E1}y'_{E2}h$

Any agent whose preferences obey (B1) through (B6) will maximize expected utility relative to a unique probability function and a utility function unique up to linear transformation.[28]

Analogously, if a preference relation $\leq$ on A satisfies (A1) through (A7), then there exist (i) a unique finitely additive, non-atomic probability function p: EE $\rightarrow$ [0, 1]; (ii) a unique increasing risk

function r: [0, 1] → [0, 1]; and (iii) a utility function unique up to linear transformation such that REU represents the preference relation ≤. If there are three such functions so that REU(f) represents the preference relation, we say that REU holds: so if ≤ satisfies (A1) through (A7), then REU holds. Furthermore, in the presence of (A2) and (A4i), if REU holds with a continuous r-function, then the remaining axioms are satisfied.

Therefore, if we assume non-degeneracy (A2) and solvability (A4i), we have:

(A1), (A3), (A4ii), (A5), (A6), (A7) are sufficient conditions for REU.

(A1), (A3), (A4ii), (A5), (A6), (A7) are necessary conditions for REU with continuous r-function.

The proof of this theorem, with references to details found in Kobberling and Wakker and in Machina and Schmeidler, can be found in my Risk and Rationality (book ms.).

6. Comonotonicity and Global Properties

In this section, I will explain how the difference between the axioms I accept and the stronger axioms the EU theorist accepts amounts to the difference between allowing agents to care about global properties – or to determine for themselves the third component of instrumental rationality – and prohibiting them from doing so.

There are roughly four types of axioms in each representation theorem. First are those that ensure that the preferences are totally ordered: (A1) and (B1). Second are those that ensure that the preferences are complex enough, but not so complex as to be unrepresentable by the real numbers: (A2) and (B2), (A4) and (B4), and (A5) and (B5). Third are those that ensure that the agent has stable views about events so that we can elicit well-defined probabilities: (A7) and (B6). Fourth, we have those that ensure that the agent has stable views for any two outcomes about the difference between including one outcome in a gamble and including the other, so that we can elicit well-defined utility differences: (A3) and (B3), and (A6) and (B6). (Utility difference ratios are the only "real fact" about utility functions, since utility functions are equivalent up to positive affine transformation.) In the axiomatizations of the two theories, axioms of the first two types are nearly identical. Axioms of the third type will not be discussed here, though I note that (B6) does double-duty in fulfilling the third and fourth functions. Axioms of the fourth type are the crux of the disagreement between my theory and EU theory. We do, however, agree about state-wise dominance (A3), and although this axiom connects values of outcomes to values of gambles, it does not play a large role in eliciting a cardinal utility function. The disagreement is really about the axioms ((A6) and (B6)) that fix utility *differences*.

The axioms of EU theory are stronger than my axioms in that I only accept weaker versions of (B5) and (B6), their comonotonic counterparts (A5) and (A6). On the face of it, the axioms of EU theory also appear weaker than my axioms in a sense, because (B3) is weaker than my (A3), (B4) is weaker than

my (A4), and I add (A7). But expected utility maximizers do obey these axioms – they follow from (B1)-(B6) – and so the standard theorist would not dispute them.[29] (Axioms (A1) through (A7) are strictly weaker than (B1) through (B6).) It is worth noting that these are essentially the axioms needed to guarantee a stable and unique probability function. Thus, we can see that the disagreement I have with the EU theorist is about whether agents are required to conform their preferences to (B5) and (B6) or only to the weaker (A5) and (A6). And since (A5) and (B5) serve primarily to ensure that utility values range over real numbers (e.g., are not infinite), and both serve that exact purpose in the presence of the respective other axioms, the substantial disagreement between us is whether rationality requires Unrestricted Tradeoff Consistency or only Comonotonic Tradeoff Consistency.

The goal of this section is to spell out how the difference between UTC and CTC maps on to the idea of being sensitive to global properties: specifically, to show that accepting a restricted versions of Tradeoff Consistency allows agents to care about global properties in the ways suggested in section 3.

Since comonotonicity will be an important concept in the discussion, let me remind the reader what it is for two acts to be comonotonic. Acts f and g are comonotonic if there are no events E_1 and E_2 such that $f(E_1) > f(E_2)$ and $g(E_1) < g(E_2)$: if E_1 leads to a better outcome than E_2 on f, then E_1 leads to at least as good an outcome as E_2 on g. Again, a comoncone is a set of gambles that are all pairwise comonotonic: all the acts in a comoncone order the events in the same way. So for each comoncone, we could order the events such that for each act in the comoncone, the agent weakly prefers events that are later in the ordering. We could, if we like, think of a comoncone as corresponding to a preference ordering over events.

Here is an example to illustrate the idea of a comoncone. A gamble that yields \$50 if a coin lands heads and \$0 if a coin lands tails is comonotonic with a gamble that yields \$100 if that same coin lands heads and \$99 if it lands tails – in either case, the agent would rather see heads than tails. A gamble that yields \$0 if the coin lands heads and \$50 if the coin lands tails is not comonotonic with either gamble, since the agent would rather see tails than heads. A 'gamble' that yields \$70 no matter how the coin comes up is comonotonic with all the gambles mentioned so far: the heads outcome is at least as good as the tails outcome, so it is comonotonic with the first two gambles mentioned, and the tails outcome is at least as good as the heads outcome, so it is comonotonic with the third gamble. The first two gambles and this 'gamble' together form (part of) a comoncone, the comonocone in which heads is weakly preferred to tails; and the third gamble and this 'gamble' form (part of) a different comoncone, the comoncone in which tails is weakly preferred to heads. So we can see that a gamble can be contained in more than one comoncone.

Recall from the previous section the idea of tradeoff equality. The relation ~*(F) holds of (xy, zw) when there are two gambles f and g contained in F (recall: F is a set of acts on some finite-event

partition) and an event E such that the agent is indifferent between the gamble that agrees with f on ¬E but yields x on E and the gamble that agrees with g on ¬E but yields y on E; and the agent is indifferent between the gamble that agrees with f on ¬E but yields z on E and the gamble that agrees with g on ¬E but yields w on E. That is, $x_Ef \sim y_Eg$ and $z_Ef \sim w_Eg$. Note that each of these four gambles must be contained in F, and that E must be non-null. Again, the idea behind tradeoff equality is that receiving x rather than y in E plays the same role as receiving z rather than w in E: they both exactly compensate for getting f rather than g in the remaining states.

What we are ultimately interested in is the utility contribution each outcome makes to each gamble it is part of: this will help us determine the utility values of outcomes. More precisely, since utility *differences* are what matter, we are interested in the utility contribution that x rather than y makes to each gamble. And tradeoff equality gives us a way to begin to determine this: if getting y rather than x in event E and getting z rather than w in event E both exactly compensate for getting f rather than g in event ~E, then they make the same difference in utility contribution in event E in those gamble pairs. In order to get from these differences in utility contributions to utility full stop, we need to fix when two pairs making the same difference in utility contribution means that they have the same difference in utility. And to do this, we will identify the conditions under which if two pairs have the same difference in utility (full stop), they must make the same difference in utility contribution, and constrain the rational agent to treat a pair consistently in these situations – to consistently make tradeoffs. Tradeoff consistency axioms provide such a constraint. Recall these axioms, from above:

Unrestricted Tradeoff Consistency (UTC): Improving an outcome in any ~*(F) relationship breaks that relationship. In other words, xy ~*(F) zw and y' > y entails ¬(xy' ~*(F) zw).

Comonotonic Tradeoff Consistency (CTC): Improving an outcome in any ~*(C) relationship breaks that relationship. In other words, xy ~*(C) zw and y' > y entails ¬(xy' ~*(C) zw).

The difference between the two axioms is that UTC says that if two tradeoffs are equal, they must be equal irrespective of the gambles they are embedded in and the event they are substituted in for; but CTC says that this only holds when restricted to gambles in the same comoncone. According to UTC, whether x-rather-than-y plays the same compensatory role as w-rather-than-z does not depend on the structure of the gambles involved. But according to CTC, it can so depend: it can depend on whether each outcome is in the same structural position in the gamble. This will become clearer shortly.

CTC follows from UTC, but not vice versa. Note that CTC is stronger than the idea that UTC must hold on every comoncone: it is stronger because if xy ~*(C_1) zw holds for some comoncone C_1, then CTC entails that for any other comoncone C_2, xy' ~*(C_2) zw cannot hold. We can also point out that for REU maximizers, xy ~*(C)zw holds just in case u(x) – u(y) = u(z) – u(w), and for EU maximizers,

xy~*(A)zw holds just in case u(x) – u(y) = u(z) – u(w). So in each theory, tradeoff equality holds when utility differences are equivalent.

Unrestricted Tradeoff Consistency entails that the utility contribution made by each outcome is **separable** from what happens in other states. In other words, y-in-E rather than x-in-E makes the same difference to the overall gamble (it exactly compensates for the same subgambles) regardless of what happens in ~E. Furthermore, y rather than x makes the same value difference regardless of which event the substitution occurs in – not in terms of absolute value, but in terms of which other tradeoffs it is equivalent to. To clarify: substituting y rather than x into a gamble will make a different value difference depending on the event the substitution occurs in, merely because the more probable the event, the bigger value difference it will make; however, if substituting y rather than x for some event in some gamble makes the same difference as substituting w rather than z for that same event in that same gamble, then for *any* event and any gamble, substituting y rather than x in that event in that gamble makes the same value differences as substituting w rather than z in that event and that gamble. This is what allows us to calculate the difference between z and w simpliciter: we can state exactly which other pair-wise differences it is equivalent to, and these pairwise equivalences will be relativized neither to a gamble nor to an event. (As mentioned, under EU theory, when xy ~*(A)zw, u(x) – u(y) = u(z) – u(w)). Therefore, the value that each outcome contributes to a gamble will be independent of ("separable from") the other outcomes that might result from the gamble.

Comonotonic Tradeoff Consistency entails that the utility contribution made by each outcome is only separable from what happens in other states if we stay within a single comoncone. In other words, y-in-E rather than x-in-E makes the same difference to the overall gamble as long as E occupies the same position in the "event ordering" in each relevant gamble. But still, if we remain in the same comoncone, then which event E is will not matter, so the value difference a trade makes will be relativized to a gamble, but not to an event. Again, under REU theory, when xy ~*(C)zw, u(x) – u(y) = u(z) – u(w).

So why would not staying within a comoncone make a different to the utility contribution that y-in-E rather than x-in-E makes? To make things concrete, let's assume we have a set of preferences that satisfies Comonotonic Tradeoff Consistency, but not Unrestricted Tradeoff Consistency. So, for example, consider four gambles f, g, h, j and five outcomes x, y, y', z, w, and consider some of the gambles we might get by making a replacement on event E: z_Ef, w_Eg, x_Ef, y_Eg, z_Eh, w_Ej, x_Eh, y'_Ej. Let's assume that the agent's preferences are as follows, where "x < f(s)" is shorthand for $(\forall s \in S)(x < f(s))$:

$z < w < j(s) < h(s) < x < y < y' < g(s) < f(s)$

$z_Ef \sim w_Eg$

$x_Ef \sim y_Eg$

$z_Eh \sim w_Ej$

$x_E h \sim y'_E j$

Note that E is the worst event (the event with the worst outcome) in the first six of these "replacement" gambles ($z_E f$, $w_E g$, $x_E f$, $y_E g$, $z_E h$, $w_E j$), and that E is the best event in the last two ($x_E h$, $y'_E j$). These preferences fail Unrestricted Tradeoff Consistency because $xy \sim^*(A)zw$ and $xy' \sim^*(A)zw$, but they don't fail Comonotonic Tradeoff Consistency because although $xy \sim^*(C)zw$, we cannot derive $xy' *(C)zw$ because $z_E h$, $w_E j$, $x_E h$, and $y'_E j$ are not in the same comoncone.

So for an agent who has these preferences, w-in-E rather than z-in-E when E is the worst event makes the same difference as y-in-E rather than x-in-E when E is the worst event. But w-in-E rather than z-in-E when E is the worst event makes the same difference as **y'**-in-E rather than x-in-E when E is the *best* event. Furthermore, y'-in-E rather than x-in-E is a better trade than y-in-E rather than x-in-E. Therefore, when E is the best event, it takes a better trade to make up for a specific bad trade than it does when E is the worst event. Put succinctly, the difference that y-in-E rather than x-in-E makes to the gamble is smaller when E is the best event rather than the worst.[30]

So why might the value contribution of y-in-E rather than x-in-E be less when E is the best event? There are two possibilities. The first is that E is considered more likely when it has a worse outcome associated with it, and less likely when it has a better outcome associated with it. In this case, the agent would not have a fixed view of the likelihood of events but would instead be pessimistic: he would consider an event less likely simply because its obtaining would be good for him. But the axiom of Strong Comparative Probability (A7) rules this interpretation out: in the presence of Machina and Schmeidler's other axioms, it entails that an agent has a stable probability distribution over events.

The second possibility is that what happens in E matters less to the agent, not because E itself is less likely, but because this feature of the overall gamble plays a smaller role in the agent's considerations. If an agent is more concerned with guaranteeing himself a higher minimum, for example, then tradeoffs that raise the minimum are going to matter more than tradeoffs that raise the maximum. I stressed that one thing an agent must determine in instrumental reasoning is the extent to which he is willing to trade off a guarantee of realizing some minimum value against the possibility of getting something of much higher value: that is, the extent to which he is willing to trade raising the minimum against raising the maximum. And, again, this is because agents must determine how to structure their goals.

So we can now see that restricting Tradeoff Consistency to gambles within the same comoncone captures the idea that agents who are risk-averse in the sense of caring about global (or structural) properties are structuring their goals differently than EU maximizers. Unrestricted Tradeoff Consistency says that trades must have the same value regardless of how they affect the structural properties of gambles. But Comonotonic Tradeoff Consistency says that the difference a trade makes depends not just

on the difference in value between the outcomes in the particular state, but on where in the structure of the gamble this difference occurs. If the agent cares about these structural properties then he will only obey the comonotonic version of the axiom.

What does Comonotonic Tradeoff Consistency rule out, then? Recall our example from above. Comonotonic Tradeoff Consistency allowed that y'-in-E rather than x-in-E when E is the best event made the same difference as w-in-E rather than z-in-E when E is the worst event. But y'-in-E rather than x-in-E when E is the best event can't make the same difference as w-in-E rather than z-in-E when E is the **best** event. More generally, the relative value of tradeoffs must be stable when the tradeoffs occur in the same structural part of the gamble: that is, when there are no global considerations at issue.

So the defender of UTC and the defender of the weaker CTC have different views on what it is to consistently value outcomes. According to the proponent of UTC, an agent consistently values outcomes if the (comparative) contribution each outcome makes to a gamble is the same regardless of the gamble. According to the proponent of CTC, an agent consistently values outcomes if the (comparative) contribution each outcome makes to a gamble is the same regardless of the gamble – as long as the outcome doesn't also figure in differently to the gamble's structural properties. Again, if preferences obey UTC, then the contribution of each outcome to the value of a gamble is **separable**: it does not depend on which other outcomes the gamble contains. If preferences obey CTC, then the contribution of each outcome to the value of a gamble is **semi-separable**: it does not depend on which other outcomes the gamble contains, unless which other outcomes a gamble contains affects the relative ranking that outcome occupies in a gamble.

7. Conclusion

I have proposed a theory on which agents subjectively determine the three elements of practical rationality: their utilities, their credences, and the tradeoffs they are willing to make in the face of risk. In this paper I have discussed how allowing agents to subjectively determine which sorts of tradeoffs they are willing to make corresponds to adopting a weaker set of axioms on preferences than those endorsed by the EU theorist. On EU theory, which tradeoffs an agent is willing to make must be determined solely by the outcomes and events those tradeoffs involve. This means that lowering the value of what happens in an event has the same effect on the value of the gamble regardless of what happens in the rest of the gamble. However, on REU theory, agents can care about where in the structure of the gamble the tradeoffs occur. Therefore, the effect on the value of the gamble can depend on whether it is the value of the minimum or maximum that is lowered. And even if the agent assigns the same probability to events E and F, she needn't think that lowering the value of E in exchange for raising the value of F (by the same utility) is an acceptable tradeoff. In particular, if she is risk-avoidant, the worst-case scenario may be

more important to her than the best-case scenario, and so this may not be an acceptable tradeoff when the prize in E is already worse than the prize in F. Now that we've seen the difference between what EU theory requires of agents and what my more permissive theory requires of them, we can properly address the question of which theory captures the requirements of practical rationality.

[1] To say that two utility functions u(x) and u'(x) are equivalent up to positive affine transformation means that there are some constants a and b where a is positive and $au(x) + b = u'(x)$.

[2] For the purposes of this paper, I will use the term "risk-averse" neutrally: an agent is risk averse with respect to some good (say, money) iff she prefers a sure-thing amount of that good to a gamble with an equivalent mathematical expectation of that good. For a more general definition of risk aversion that is compatible with what I say here and that captures the idea that a risk-averse person prefers a gamble that is less spread out, see M. Rothschild and J. Stiglitz (1970), "Increasing Risk: I. A Definition," *Journal of Economic Theory.*

[3] In all these examples, I will assume that probabilities are given, to simplify the discussion. But the probabilities involved should be assumed to be the agent's subjective probabilities.

[4] Particularly clear expositions of this view appear in the following: (1) Patrick Maher (1993), Betting on Theories, Cambridge: Cambridge University Press; (2) John Broome (1999), "Utility," in Ethics out of Economics, Port Chester, NY, USA: Cambridge University Press; (3) James Dreier (2004), "Decision Theory and Morality," Chapter 9 of Oxford Handbook of Rationality, eds. Alfred R. Mele and Piers Rawling, Oxford University Press.

[5] Matthew Rabin (2000), "Risk Aversion and Expected Utility Theory: A Calibration Theorem," *Econometrica*, pg. 1282.

[6] Rabin states his results in terms of changes from initial wealth levels because he thinks that the correct explanation for people's risk aversion in modest stakes is *loss aversion* of the kind discussed in Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47, pg. 263-291..

[7] Example due to Maurice Allais (1953), "Criticisms of the postulates and axioms of the American School," reprinted in Rationality in Action: Contemporary Approaches, Paul K. Moser, ed., Cambridge University Press, 1990. Amounts of money used in the presentation of this paradox vary.

[8] For if L_1 is preferred to L_2, then we have $0.1(u(\$5m)) + 0.9(u(\$0)) > 0.11(u(\$1m)) + 0.89(u(\$0))$. Equivalently, $0.1(u(\$5m)) + 0.01(u(\$0)) > 0.11(u(\$1m))$. And if L_4 is preferred to L_3, then we have $u(\$1m) > 0.89(u(\$1m)) + 0.1(u(\$5m)) + 0.01(u(\$0))$. Equivalently, $0.11(u(\$1m)) > 0.1(u(\$5m)) + 0.01(u(\$0))$. These two contradict; so there is no utility assignment that allows for the common Allais preferences.

[9] This talk may faze a certain kind of constructivist. We could recast it in terms that are acceptable to the constructivist as follows. If risk-preferences are based only on local considerations so that the agent obeys the axioms of EU theory, then the utility function as determined by EU theory will reflect these even if it doesn't correspond to anything 'real.' If risk-preferences are based on both kinds of considerations so that the agent doesn't obey the axioms of EU theory, then constructivist EU theory will read the agent as not having a utility function.

[10] If contra our supposition, $u(O_2) \leq u(O_1)$, then the value of the gamble would be $r(p(\sim E))u(O_1) + (1 - r(p(\sim E))u(O_2)$, i.e. $r(1 - p(E))u(O_1) + (1 - r(1 - p(E))u(O_2)$, which is not necessarily equivalent to $r(p(E))u(O_2) + (1 - r(p(E))u(O_1)$.

[11] For further discussion of this point, see my book manuscript.
[12] $L_1 > L_2 \Leftrightarrow r(.1)[u(\$5m) - u(\$0)] > r(.11)[u(\$1m) - u(\$0)]$.
$L_4 > L_3 \Leftrightarrow (1 - r(.99))[u(\$1m) - u(\$0)] > r(.1)[u(\$5m) - u(\$1m)]$.
These inequalities hold jointly only if $r(0.11) - r(0.1) < 1 - r(0.99)$.
[13] David Schmeidler (1989), "Subjective Probability and Expected Utility without Additivity," *Econometrica* 57, pp 571-587. Itzhak Gilboa (1987), "Expected Utility with Purely Subjective Non-Additive Probabilities." *Journal of Mathematical Economics* 16: 65-88.
[14] John Quiggin (1982), "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization* 3, pg. 323-343.
[15] Schmeidler's (1989) version includes some objective probabilities to derive the decision weights.
[16] Given the similarity in the formalism, I could present REU theory as a generalization of AU theory to decision making with subjective probabilities: i.e., as "subjective" anticipated utility theory. But to do so would be misleading about what the functions in Quiggin's theory are meant to refer to: on Quiggin's theory, the decision weights are themselves subjective probabilities. Therefore, AU interprets decision makers as having credences that are different from their known objective probabilities (AU maximizers are usually referred to as "optimistic" or "pessimistic"). This interpretation makes agents who are sensitive to global properties automatically irrational, since their credences are different from the objective probabilities and their credence in each event is not independent of which outcome occurs on that event. On the contrary, on my theory there is room for agents to believe a coin is fair even if the heads outcome does not get a decision weight of 0.5. Furthermore, the philosophical foundations of my theory – the way in which I analyze means-ends reasoning – are very different from Quiggin's. In addition, the formalism itself is a fairly intuitive generalization of EU theory: as Peter Wakker notes, Allais himself proposed, but rejected, this possibility. Therefore, I think it is best to view my theory as somewhat similar in formalism, but different in philosophical commitments, to the rank-dependent theories present in the literature.
[17] See Leonard Savage (original 1954, second edition 1972), The Foundations of Statistics, New York: Dover Publications, Inc. See also Frank P. Ramsey (1926), "Truth and Probability," in Ramsey, 1931, The Foundations of Mathematics and other Logical Essays, Ch. VII, pp 156-198, edited by R.B. Braithwaite, London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company. For a survey of representation theorems for EU theory, see Peter Fishburn (1981), "Subjective Expected Utility: A Review of Normative Theories," *Theory and Decision* 13. A different sort of representation theorem is due to Jeffrey, Richard (1965), The Logic of Decision, McGraw Hill. For Jeffrey, the state space and the outcome space are the same, and each outcome is a gamble over other outcomes, i.e., there are no "final outcomes." As a result, his uniqueness result for the utility function is weaker.
[18] Note that the utility function is only unique up to positive affine transformation. Therefore, only the facts that are common to all of the utility functions, e.g., the relative size of the utility intervals between outcomes, are rightly called facts about the agent's utilities.
[19] See my book manuscript.
[20] Veronika Kobberling and Peter Wakker (2003), "Preference Foundations for Non-expected Utility: A Generalized and Simplified Technique," *Mathematics of Operations Research* 28, pp. 395-423. Mark J. Machina and David Schmeidler (1992), "A More Robust Definition of Subjective Probability," *Econometrica* 60(4).
[21] In the denotation of the spaces, I follow Machina and Schmeidler (1992).
[22] Machina and Schmeidler (1992), p. 749.
[23] Kobberling and Wakker (2003), p. 400. On pg. 403, Kobberling and Wakker point out that we can also define a comoncone on an infinite state space, although this is not necessary for our purposes.
[24] Kobberling and Wakker (2003), p. 398. The concept of a standard sequence, however, does not originate with them.
[25] Id, pg. 396-397.
[26] Kobberling and Wakker (2003), p. 397.
[27] Kobberling and Wakker (2003). I change their presentation slightly, for readability.
[28] Theorem 5 in Kobberling and Wakker (2003). Note that I added non-degeneracy; they list the results of the degenerate cases (with non-unique probability and utility functions) separately; see pg. 399.
[29] We can see that EU maximizers obey them by noting that REU maximizers with a continuous r-function obey them and that EU maximization is a special case of REU maximization with a continuous r-function.
[30] Note that the agent would have $x_E h > y_E j$, by state-wise dominance ($y'_E h > y_E j$) and ordering.