Risk Aversion and Rationality

Lara Buchak, January 2010

0 Introduction

Ralph has the opportunity to participate in two gambles. In the first, a referee will flip a coin, and if it lands heads, Ralph will receive a rare Elvis stamp (Ralph is a stamp collector). In the second, the referee will flip a different coin, and if it lands tails, Ralph will receive a nice pair of gloves. Ralph believes both coins to be fair. Now a trickster comes along, and offers Ralph a sort of insurance: for a few cents, she will rig the game so that the first coin determines both outcomes – if it lands heads, he gets the Elvis stamp, and if it lands tails, he gets the gloves – and therefore that Ralph is guaranteed to receive some prize. Ralph values the two goods *independently* in the sense that having one does not add to or decrease from the value of having the other; they are not like, say, a left-hand glove and a right-hand glove. Receiving a prize does not have any value apart from the value of the prizes themselves: Ralph does not like winning for its own sake. And he is not the sort of person who experiences regret or disappointment when he might have gotten something but didn't; he only cares about what he actually has. He decides that the trickster's deal is worthwhile – it would be nice to guarantee that he gets something no matter what – so he decides to pay a few cents to rig the game. We can represent his options schematically as follows:

	HH	HT	TH	TT
Deal 1	*Elvis stamp*	*Elvis stamp and gloves*	*Nothing*	*Gloves*
Deal 2	*Elvis stamp*	*Elvis stamp*	*Gloves*	*Gloves*

Ralph prefers deal 2 to deal 1. This seems very reasonable. Many of us would have a similar preference. And yet, standard decision theory rules this out: it cannot represent his preferences, and therefore judges Ralph to be irrational.[1]

Margaret is very clear on how she values small amounts of money. Receiving $50 is just the same to her whether she starts with $0 or $50, and she feels similarly about all small increments of money. We might say that she values money *linearly*: every dollar that she receives is worth as much to her as the previous ones, at least for amounts of money less than, say, $200. And Margaret is equally clear on how she values bets: she prefers $50 to a coin flip

[1] He is irrational if he has any preference (besides indifference) between these two deals. Note that this is a general schema for a counterexample to the theory, so if the reader is not motivated by the example with these goods, he can substitute in some other goods: for example, a new house and a prestigious non-monetary philosophy award. Of course, the above qualifications – that the goods are independent and that the agent is not the sort who experiences regret or likes winning for its own sake – are essential. Sections of this paper will be devoted to showing that these assumptions are plausible.

between \$0 and \$100. If she takes the former, she will certainly get \$50, and the possibility of getting \$100 is not enough to make up for the (equally likely) possibility of getting \$0 – she would rather guarantee herself \$50 than take that chance. Again, these preferences seem appealing to many people, and are at least understandable. But standard decision theory cannot represent Margaret's preferences, and judges her – like Ralph – to be irrational.

Finally, in a classic example due to Maurice Allais, commonly known as the Allais paradox, people are presented with a choice between L_1 and L_2 and a choice between L_3 and L_4, where the gambles are as follows:

L_1: \$5,000,000 with probability 0.1, \$0 otherwise.
L_2: \$1,000,000 with probability 0.11, \$0 otherwise.
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.

People tend to choose L_1 over L_2, and L_4 over L_3: in the first pair, the minimum amount that one stands to walk away with is the same for either gamble, and there is not much difference in one's chances of winning *some* money – but L_1 yields higher winnings; in the second pair, however, the minimum amount that L_4 yields is a great deal higher than the minimum that L_3 yields.[2] Again, these preferences are understandable (most people express them), but standard decision theory cannot accommodate them, and, again, must judge them to be irrational.

In this paper, I defend the rationality of certain preferences – like Ralph's, Margaret's, and the Allais choosers' – that decision theory treats as irrational, and I offer a more permissive theory of rational decision making. What the preferences I defend have in common is that they all stem from the decision maker's attitude towards *risk*. The standard theory (expected utility theory) rules out caring about how the possible outcomes in a gamble relate to each other: for example, about the best or worst prize an agent might receive, about the difference in value between these two, about the variance among outcomes, or about the proportion of outcomes reaching a certain threshold. I show that we get an interesting and more intuitive version of utility theory if we relax the assumption that rules out these preferences, and I show that three classic arguments against these preferences fail: contra the standard theory, these preferences are in fact rational.

The first section is devoted to explicating the standard theory and the constraints it places on how agents treat risk. I show that the decision-theoretic notion of risk aversion is not the commonsense one, and that this leads to decision theory's failure to capture some intuitive preferences – though I do not yet argue that these preferences are rational. Along the way, I will

[2] Allais (1953), pg. 132. Amounts of money used in the presentation of this paradox vary.

flesh out the assumptions in the examples mentioned. I then formalize what I take to be the more intuitive way to think about risk aversion. In my preferred theory, possible outcomes combine to yield the value of a gamble in a way that is sensitive not just to what happens in each state, but also to "global" properties of gambles, such as the worst possible outcome and the spread of possible outcomes, that contribute to what is more naturally called the riskiness of a gamble.

The main difference between my theory and the standard theory is that while the standard theory has two parameters that determine an agent's preferences – a subjective utility function and a subjective probability function – mine adds an additional parameter: a subjective risk function. Furthermore, agents who conform to my theory frequently violate an axiom that expected utility maximizers do not violate, known as the sure-thing principle (hereafter, STP). I am not the first to propose a theory that violates STP, but to my knowledge I am the first to propose a non-expected utility theory in which agents have both subjective probabilities and subjective risk attitudes (on some theories, for example, probabilities are given, and on others, risk-attitudes *take the place of* subjective probabilities)[3]; I also differ from most proponents of non-expected utility theories in arguing that my theory describes *rational* preferences, not deviations from rationality. The remaining three sections are devoted to arguing just this: I explore whether STP, or expected utility maximization more generally, is a requirement on rational preferences; if it is, then it will be irrational for agents to care about risk in the way I describe. So my theory may be of technical and psychological interest, but it will be of no interest to agents trying to determine what they should do, or to theorists trying to determine which sets of preferences are rational or trying to explain rational behavior. I consider three arguments purporting to show that agents are rationally required to obey STP and maximize expected utility:[4] that STP follows from an undeniable intuition, that agents who do not maximize expected utility do worse over the long run, and that it is conceptually impossible for rational agents to violate STP because the (rational) reasons for their preferences render purported violations not violations at all. I claim that none of these arguments succeeds. I conclude that there are attitudes towards risk that give rise to preferences that are not capturable by standard decision theory, but are nonetheless rational.

---

[3] Machina (1983 and Machina (1987) contain an excellent discussion of some of the non-expected utility theories (in particular, theories that use objective probabilities) that have been proposed in response to particular psychological findings. See also Kahneman and Tverksy (1979), Quiggin (1982), and Schmeidler (1989).

[4] Obeying STP and maximizing expected utility are equivalent in the presence of the other axioms of expected utility theory, which I take to be non-controversial for the purposes of this paper. The first and third argument are specifically a defense of STP, and the second of EU theory in general (that is, of the conjunction of all the axioms).

First, a note about terminology: I will use the term *option* to refer to the things amongst which the decision maker must choose. I will use the term *outcome* to refer to any of the final results of an option. *States* are the various contingencies which might obtain. A *gamble* is a function from states to outcomes; that is, a gamble specifies which outcome obtains in each possible state of the world – it might be the same outcome in every state, or different outcomes in some states than in others. As an example, consider an agent deciding between two options: not bringing an umbrella and bringing an umbrella. The relevant states are "it rains" and "it does not rain" and the outcomes are (O_1) not carrying an umbrella and getting wet, (O_2) not carrying an umbrella and not getting wet, and (O_3) carrying an umbrella and not getting wet. The option "no umbrella" can be thought of as the gamble that yields O_1 if it rains and O_2 if it does not rain, and the option "umbrella" as the 'gamble' that yields O_3 either way. We can represent his decision problem in a chart:

	Rain	No rain
No umbrella	Not carry, wet	Not carry, dry
Umbrella	Carry, dry	Carry, dry

In general, we will list the possible states across the top row, the decision maker's options in the first column of each row, and the gambles corresponding to each option across each row (with each outcome listed in the column corresponding to the state in which it is the final result of the gamble on that row). The distinction between states and outcomes does not matter for our purposes. However, in making a distinction between gambles and outcomes, I am following Savage rather than Jeffrey.[5] I leave it open what outcomes are – I'll sometimes speak of receiving a prize, and sometimes of making a proposition true (e.g. the proposition that Ralph gets a pair of gloves). It should be obvious when I'm using each, and it makes no difference to my argument; the main thing to note is that the inputs of the utility function and the preference relation are generally taken to be total worlds rather than individual goods: e.g., the claim that an agent has a preference for a pair of gloves rather than $50 is shorthand for the claim that the agent prefers W_g, the world which has all the features of the actual world and the agent has an additional pair of gloves, to $W_{\$50}$, the world which has all the features of the actual world and the agent has an additional $50; and "u(gloves)" is shorthand for "u(W_g)."

[5] Savage (1954), pp 13-17. I use the term *gambles*, whereas Savage uses the term *acts*; the differences are not important. Jeffrey (1965), pp 145-150. It might be that the formalism of my theory can ultimately do without specifying beforehand what counts as an outcome and what counts as a gamble, but I certainly need the distinction between outcomes and gambles for the arguments of this paper to be tractable. I take it the distinction is at least intuitive.

1. Risk, intuitively

In expected utility (EU) theory, the structure of people's preferences is fixed: as long as a decision maker's preferences obey certain basic axioms,[6] there is some way to assign values to outcomes (a numerical "utility" function of outcomes) such that between any two options, the agent prefers the option whose expected utility is higher. We say that a utility function *represents* an agent under expected utility theory just in case for all gambles X and Y, the agent weakly prefers X to Y iff $EU(Y) \leq EU(X)$.[7] If a decision maker's preferences obey the axioms of expected utility theory, then not only can we derive a utility function that represents his preferences, that utility function will be unique up to positive linear transformation.[8] Therefore, the values of all the possible outcomes – the utility function – are determined by plugging an agent's ordinal preferences into that structure.

There are two pictures of what the utility function is: in Bengt Hansson's terminology, the *realistic* interpretation and the *formalistic* interpretation.[9] On the realistic picture, the utility function has a meaningful empirical interpretation: it is a measure of how much an agent desires something or the degree of satisfaction that having it could bring him; that is, it represents some pre-existing value that the agent has. On this picture, the agent may have intuitions about whether a particular utility function accurately captures his values. On the formalistic picture, which seems to have more widespread endorsement among contemporary philosophers,[10] utility is not meant to be a measure of goodness, or of any quantity in the head or in the world; rather, it is just whatever quantity plays the correct role in the theory: the quantity whose mathematical expectation an agent maximizes. It may, for example, incorporate the agent's actual values and how he aggregates them. Thus, on this picture, we cannot have intuitions about (cardinal) utility

---

[6] Including, among others, completeness, transitivity and STP.

[7] Actually, this needn't be strictly true: if an agent occasionally deviates from what the function entails because of an inability to make precise calculations, or because of a brief lapse in judgment, then we can say that the function represents him because it represents his ideal preferences (or approximates his actual preferences) – but there cannot be *widespread* deviation that an agent would *endorse* even at his most clearheaded.

[8] That this is possible for agents who meet the axioms is shown by various representation theorems. E.g. Savage (1954; 1972). Savage's axiomatization, like that of Ramsey (1926) and von Neumann and Morgenstern (originally in von Neumann and Morgenstern (1944); described in Resnik (1987)), yields a utility function that is unique up to positive linear transformation. The uniqueness condition for Jeffrey's (1965) axiomatization is more complex.

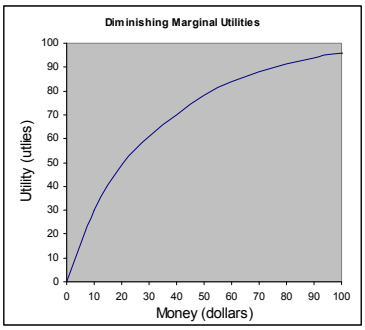
[9] Hansson (1988), pp 142-146.

[10] Particularly clear expositions of this view appear in Maher (1993), Broome (1999), and Dreier (2004). Hansson (1988) guesses that the realistic interpretation is more common in the literature – he attributes this view to Jeffrey (1965), among others – and thinks that it is certainly the historically correct interpretation. Perhaps this difference in how we see the field can be explained by the increasing popularity of the formalistic interpretation in recent years. In any case, both interpretations are philosophically important, and, crucially, both must deal with the problem I discuss.

values. On either picture, if an agent (or an idealized or approximated version of that agent) is not representable by a utility function, then he is considered irrational – or incoherent.

As I mentioned, for agents who are representable, any variation in preferences over gambles among these agents is going to show up as variation in utility values of the outcomes,[11] since the value of a gamble is simply the mathematical expectation of its outcomes. And the fact that some agent does not like to take risks – and has preferences that reflect this – is going to show up as her having a particular utility function; again, it is going to show up in the values her utility function assigns to *outcomes*.

For example, many people's preferences display risk aversion in the following sense: an individual would rather have \$50 than a 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.[12] And when these preferences are plugged into the structure of EU theory, we get that utility is a function of monetary value, but one that diminishes marginally: as the size of one's winnings increases, additional amounts of money add less value. If an agent is going to get \$50, then getting an extra \$50 doesn't increase the value of the prospect as much as it does when he is initially not going to get anything. Thus, preferences that are risk averse with respect to money imply, on the standard theory, a concave utility function.



And vice versa: when a utility function is concave, a risky gamble (say, the coin flip between \$0 and \$100) will always have an expected utility that is less than the utility of its expected dollar value (\$50). This is because the difference between the utility of \$0 and the utility of \$50 is larger than the difference between the utility of \$50 and the utility of \$100, so the

[11] To simplify the discussion in this section, we assume that the probabilities of the outcomes in the gambles are fixed and known to the agents.

[12] 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.

value of the higher outcome does not entirely make up for the value of the lower outcome: the higher outcome is not better than $50 by as much as the lower outcome is worse than $50.[13]

On the standard theory, then, aversion to risk is equivalent to diminishing marginal utility. What this means depends on which conception of utility we adopt: if we adopt the realistic interpretation, then diminishing marginal utility is the psychological explanation for risk aversion. On the other hand, if we adopt the formalistic interpretation, then the concavity of the utility function needn't correspond to a property of the agent's introspectively felt values. So it won't be a mark against the supposed equivalence of risk aversion to diminishing marginal utility that the agent himself doesn't feel that his utility values do diminish. However, I will show that a failure to account for all of the psychological phenomena that underlie risk averse preferences creates a problem for expected utility theory on both interpretations. I will proceed by pointing out the problem for the realistic interpretation, and then show that the formalistic interpretation does not avoid this problem.[14]

On the realistic interpretation, we might call the diminishing marginal utility explanation of risk aversion *saturation*: one's preferences display risk aversion because money loses its value as one has more of it. However, there is another, very different psychological phenomenon that might give rise to risk-averse behavior.

Consider two people who collect stamps. Alice is only interested in obtaining one Elvis stamp (she likes her collection to be diverse): once she has one, a second Elvis stamp is next to worthless. Bob, on the other hand, has an insatiable appetite for Elvis stamps (he doesn't get "saturated" with respect to Elvis stamps), but he does not like to take risks. In general, he thinks that the *possibility* of getting something good doesn't make up for the *possibility* of getting something bad (when the bad thing is as bad as the good thing is good, and there is an equal chance of each). He might even dislike taking risks so much that he always just wants the gamble with the highest minimum. When offered a choice between one Elvis stamp on the one hand, and a coin flip between two Elvis stamps and none on the other hand, both Alice and Bob will choose the latter. But they have very different reasons for this preference: Alice values outcomes in a

---

[13] To keep things simple, I am talking about (fair) coin flips, where an agent has the same chance of getting each outcome. But my remarks are intended to be general: the value of the higher outcome (weighted by the probability of getting it) does not make up for the value of the lower outcome (weighted by the probability of getting it). If utility diminishes marginally, then if we consider the sure-thing $x compared to a gamble whose mean value is $x, then the agent will prefer the former: the higher possible values of the latter gamble do not raise the mean utility of that gamble by as much as the lower values lower it.

[14] Hansson (1988) also argues that EU theory rules out explicitly considering risk when making decisions, on either interpretation. I will discuss an example of his shortly.

particular way (she quickly becomes saturated with respect to Elvis stamps), while Bob cares about how outcomes of particular value are arranged across the possibility space.[15]

Similarly, there are two different attitudes that could lead to a preference for $50 rather than a coin-flip between $0 and $100. It might be that an increase from $0 to $50 is worth more than an increase from $50 to $100. But it equally might be that a person really does value money linearly (as Margaret claims to), but when she evaluates the attractiveness of a *gamble*, she does not just care about the final dollar amounts she might receive, in isolation – she also cares about how these amounts are arranged across the possibility space, since she does not yet know which possibility will obtain. In other words, she cares not just about the components of a gamble, but about global features of the gamble, e.g., its minimum value, its maximum value, the interval between these two values, the variance, and so forth. The possibility that she might get the higher outcome does not entirely make up for the possibility that she might get the lower outcome: adding a good possibility (with some specific probability) to the mix does not make up for adding a bad possibility (with some specific probability) to the mix, or for lowering the minimum value of the gamble, even though the mean remains unchanged. I will call this explanation of risk aversion *global sensitivity*.

To illustrate my point that Ralph and Margaret care about global features of gambles, we can consider Ralph's choice as a choice about which of the (equally likely) states he would like to enhance with the possibility of a pair of gloves, when presented with the following initial setup:

	HH	HT	TH	TT
Deal	*Elvis stamp*	*Elvis stamp*	*Nothing*	*Gloves*

If the possibility of gloves is added to the HT state, we get deal 1, and if it is added to the TH state, we get deal 2:

	HH	HT	TH	TT
Deal 1	*Elvis stamp*	*Elvis stamp and gloves*	*Nothing*	*Gloves*
Deal 2	*Elvis stamp*	*Elvis stamp*	*Gloves*	*Gloves*

Since Ralph prefers deal 2, he prefers that the gloves be added to the TH state. Thus, he has a preference about how values are arranged across the possibility space, given that the value of the goods does not depend on whether he has the stamp and so the value added to each state is the same. He would like to add the gloves to a state in such a way that he increases the minimum he

[15] Note that saturation need not always give rise to risk-averse behavior. For example, we might have a person who only slightly prefers two Elvis stamps to one, but who always prefers to take the gamble with the highest maximum. This person will prefer the coin flip to the certain Elvis stamp, and thus display preferences that are "risk-seeking."

might receive from the gamble. Similarly, we can consider Margaret as faced with an initial gamble that yields $0 if a fair coin lands heads and $50 if it lands tails. Her choice is whether to add $50 to the heads state or to the tails state – and if she would be adding the same value to the state in either case (that is, if her utility function does not diminish marginally), but prefers to add it to the heads state, then she cares about how the values are arranged across the possible states.

The global sensitivity explanation says that even holding fixed how much one likes various outcomes, and thus holding fixed the average value of a gamble, which outcomes constitute the gamble and how they are arranged matter. That is, properties of gambles that do not supervene on any particular outcome or chance of getting that outcome – global properties of gambles – matter. This does not rule out marginal diminishment in the value of goods. People who have diminishing marginal utility functions can still be sensitive to global properties; I suspect that most people's utility functions do diminish marginally for very large quantities of money, and yet that many people are globally sensitive.[16] So, whereas the expected utility theorist claims that risk averse preferences always entail a diminishing marginal utility function, I claim that they might instead arise from global sensitivity (or from some combination of the two).

What my explanation does is separate the two dimensions of evaluation (of options) that get folded together by the standard theory: (1) how much an agent values certain ends and (2) how effective he rates various means of arriving at these ends to be.[17] This isn't exactly the distinction between how an agent values ends and how an agent *values* means. I am not pointing out, for example, that of two agents who share the same ends, one may think that hard work is a better way to his ends while the other thinks that schmoozing is a better way because of the intrinsic values of these means. Indeed, since ends in this case are total worlds, the value of getting good things through hard work and the value of getting them through schmoozing will be accounted for in the value the agent assigns to ends, because total worlds include facts about how an agent obtained his ends. Instead, when I talk about the agent rating the effectiveness of various means, I am talking about determining which means he thinks will most effectively achieve his ends, with only the goal of achieving those ends in mind. What I am pointing out is

[16] So far, I have not said how a person can have diminishing marginal *utilities* without obeying the standard axioms of decision theory, since utility functions are defined only for people who obey these axioms. What I mean to suggest is that the following two psychological facts can hold of an agent simultaneously: the value of goods to an agent can diminish marginally and global properties can matter to how she evaluates gambles. To show that we can define a utility function for someone who is cares about risk in the manner I suggest, we need a representation theorem. For this theorem, see my Risk and Rationality (ms).

[17] Agents also evaluate the subjective probabilities of states; this might be thought of as a third dimension of evaluation. However, this does not make a difference to my arguments: as stated, for the purposes of this paper, we can take the probabilities to be known or already determined by the agent. I claim that two agents who assign the same values to outcomes and the same probabilities to states may rate the same gamble differently in terms of its effectiveness of realizing their ends.

this: two agents could attach the very same values to certain ends (various sums of money, say). And yet, one agent might think that he can more effectively achieve his ends of getting as much money as he can by taking a gamble that has a small chance of a very high payoff, whereas the other might think that he can more effectively achieve *these same ends* by taking a gamble with a high chance of a moderate payoff. In other words, they may have different ways of aggregating value across the possibility space.

Expected utility theory claims that as long as the average values of the prizes in two gambles are the same, rationality requires that an agent does not care how the possibilities are arranged to yield that average: he must be indifferent as to whether all the possible value is concentrated in an outcome with small probability or whether the value is spread evenly over states. Another way of explaining the constraint that expected utility theory places on evaluating gambles is as follows: we can think of a gamble as a probability distribution over utility values. This distribution will have many features: a minimum, a mean value, and a variance, to name a few. According to expected utility theory, the *only* feature of this distribution that can (rationally) matter when evaluating a gamble is its mean value (i.e. its expected value); we are meant to throw out any other information as useless to determining the value of the gamble.[18] We will find preferences that violate the theory when we find agents who are sensitive to some of these other features of distributions. And since decision theory is supposed to represent all rational preferences, all of these globally sensitive agents are deemed irrational on the standard theory.

So, as mentioned, a rational agent cannot value money linearly and still prefer $50 to a coin flip between $0 and $100. Nor can Ralph, if he is rational, prefer deal 2 to deal 1, if my assumptions in the case are correct. I will spell out one of these assumptions now: that the values of the outcomes (specifically, an Elvis stamp and gloves) are independent. I mentioned that having one does not add to or decrease from the value of having the other; that is, receiving A is just as valuable whether or not Ralph already has B, and vice versa. This is a psychological fact about what Ralph values, and on the realistic interpretation of utility, it should be taken into account directly by the utility function. To capture this fact in utility terms, two goods are independent if the values of each prize individually sum to the value of both prizes together: u(A & ~B) + u(B & ~A) = u(A & B).[19] Examples of goods that are obviously not independent for most people include a left-hand glove and a right-hand glove: the right-hand glove is more

[18] This way of putting the point was suggested to me by Branden Fitelson.

[19] That is, u(A & B) – u(~A & B) = u(A & ~B) – u(~A & ~B) : the difference the presence of A makes is the same whether or not B is present. This is equivalent to u(A & ~B) + u(B & ~A) = u(A & B) + u(~A & ~B), which is equivalent to the above when we set the utility of the status quo (in which the agent has neither A nor B) to 0.

valuable to me if I have the left-hand glove, and vice versa.[20] An Elvis stamp and a different Elvis stamp are also not independent for Alice in the above example: one is less valuable to her if she already has the other.[21] As a special case of outcome independence, if the utility of money is taken to be linear in small amounts, then receiving one dollar is valued independently of receiving another dollar: u($1) + u($1) = u($2). So the psychological fact that Margaret or values small increases in her wealth independently of what other small amounts of money she has corresponds to the mathematical fact (again, if utility is interpreted realistically) that her utility function is linear in dollar amounts. Similarly, if the goods in Ralph's decision problem are independent, then the utility of deal 1 must be the utility of deal 2.[22]

The two examples discussed so far rely on the assumption that agents have some introspective access to how they value goods: in particular, an agent knows of some goods that he values them independently. As I mentioned, though, many philosophers interpret utility in the formalistic sense. Therefore, the assumption that the agent's utility function is linear is a non-starter. However, if there really are two distinct psychological phenomena – saturation and global sensitivity – then an agent who is globally sensitive but does not experience saturation will have preferences that are inconsistent with his maximizing expected utility with respect to *any* utility function. I began with the examples of Ralph and Margaret because they seem to be to be particularly clean cases illustrating the distinct psychological phenomena, but I now explain two examples that more clearly do not depend on assumptions about the utility function.

The first is adapted from Hansson's elucidation of John Watkins's point that expected utility maximization rules out taking certain kinds of risk-related considerations into account: Hansson shows that Watkins's point holds even when we interpret utility formalistically.[23] Consider a person who adopts the following two reasonable-seeming rules about taking gambles:

---

[20] Spelling this out in utility terms: u(r & ~l) and u(l & ~r) are each very small, so u(r & ~l) + u(l & ~r) < u(l & r).

[21] u(E1 & ~E2) and u(E2 & ~E1) are each almost as large as u(E1 & E2), so u(E1 & ~E2) + u(E2 & ~E1) > u(E1 & E2).

[22] Since the results of the coin flips are probabilistically independent, each state has probability ¼. So:
EU(deal 1) = ¼u(stamp and gloves) + ¼u(stamp) + ¼u(gloves)
= ¼[u(stamp) + u(gloves)] + ¼u(stamp) + ¼u(gloves) if the stamp and gloves are independent
= ½u(stamp) + ½u(gloves)
= EU(deal 2)
That the expected utilities of the two deals are equivalent is perhaps easier to see if we evaluate the expected utility of the first coin flip, and then add it to the expected utility of the second. However, I wanted to flag exactly where the assumption that the goods were independent entered in: we can only evaluate the utility of two gambles by adding their utilities if they are independent in the relevant sense anyway.

[23] Hansson (1988), pp 147-151. Watkins (1977).

R1. Whenever faced with a lottery with a 50% chance of one sum of money and a 50% chance of another which is $30 larger, be indifferent between taking that lottery and receiving the lesser prize plus $10 for certain.

R2. Whenever faced with a lottery with a 50% chance of nothing and a 50% chance of some particular sum, be indifferent between taking this lottery and receiving one-third of that sum for certain.[24]

Hansson shows that if an agent's preferences obey both of these rules, then there will be no utility function that represents the agent's preferences under expected utility maximization. The formal reason for this is that if we assume an agent is an expected utility maximizer, then R1 and R2 each imply that the utility function diminishes marginally, but they each imply a different rate of diminishment. Intuitively, though, these rules do in fact seem quite consistent with each other, and seem to describe a consistent way of valuing gambles: a 50% chance of getting $x more than some worst possible value makes a gamble worth $x/3 more than its worst possible value. Again, EU theory dictates that an agent cannot value gambles in a way that is sensitive to their minimum values in particular.

The final example that does not rely on a realistic interpretation of utility was mentioned above: the so-called Allais paradox. Like the previous example, Allais's example does not require any background information about how the outcomes are related in order to get a contradiction with expected utility theory; indeed, it relies only on the agent having two specific preferences. All we need to know is that a decision maker prefers L_1 to L_2 and L_4 to L_3. There is no assignment of utility values to the outcomes $0, $1m, and $5m such that $EU(L_1) > EU(L_2)$ and $EU(L_3) < EU(L_4)$; therefore, anyone who has these preferences violates expected utility theory.[25] Consequently, anyone who has the standard Allais preferences must be responding to properties that cannot be captured by an expectational utility function.

Expected utility theory, whether utility is interpreted realistically or formalistically, rules out caring about certain global properties of gambles. Because of this, expected utility theory rules out some preferences that seem at least prima facie reasonable. This is (I claim) because the standard theory folds together how much an agent values particular ends and how effective he considers various ways of arriving at these ends. Let us now turn to an alternative theory that allows for a wider range of attitudes towards risky gambles to count as rational.

---

[24] Hansson, pp 149-151. Language altered slightly.

[25] 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.

2 Formal representation of the two dimensions of evaluation

I've pointed out that in expected utility theory, utility functions vary from agent to agent, but utilities and probabilities interact in a set way. The utility of a gamble is its mathematical expectation: the utility of each possible outcome weighted by the probability of obtaining that outcome. So, the utility of a gamble between A and B, where the agent has a probability *p* of getting A and 1 – *p* of getting B, is p(u(A)) + (1 – p)(u(B)). This is equivalent to u(B) + p[u(A) – u(B)]. Taking B to be the less (or equally) desirable option, this latter formulation says that the value of a gamble will be the minimum value it guarantees plus the amount by which the agent might do better, scaled by the probability of doing that much better. So the possibility of getting a better outcome will increase the value of a gamble above its minimum value in a set way.

Thus, agents attach values to various outcomes, but probabilities are still of set significance: e.g., if the probability of getting a good outcome (instead of nothing) is doubled, the value of a gamble (above the status quo) is doubled, even when the new probability is 1 – that is, even when doubling the probability removes the riskiness from the gamble. As mentioned above, the value of a gamble that yields a good outcome (rather than nothing) with probability 0.5 must be half the value of that good outcome.[26] Furthermore, once two decision makers agree on the values of various outcomes, 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 are more cautious than others. 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 a coin flip between $0 and $100; the fact that there is only a 50% chance of winning the better prize may make the gamble less attractive to one person than to the other. In other words, aside from having different attitudes towards outcomes, two agents might have different attitudes towards potential ways of obtaining some of these outcomes. And to repeat, these attitudes are not most plausibly read as 'actually' having different attitudes about the value differences between $0 and $50 and between $50 and $100. Rather, we should think the agents have different ways of aggregating possible value.

One way to conceive of these differences is to read decision makers who are sensitive to global properties of gambles as weighting the interval by which they could improve over the minimum by different amounts. For some, the possibility (with some specific probability) of improving over the minimum by some amount might not count heavily in the evaluation of the gamble – on the contrary, the minimum they are guaranteed to receive will be more important. These people will display risk averse behavior: if two gambles yield the same monetary value on

[26] Relative to the status quo.

average, but one yields a specific amount no matter what (and therefore has a higher minimum), they will prefer the latter. Indeed, some individuals may be so averse to risk that they simply disregard any possibility of getting higher than the minimum value of a gamble: for these individuals, the worth of a gamble may simply be its minimum value.[27] For others, the possibility of improving over the minimum will count for a lot – and the maximum they might receive will count heavily in the evaluation of the gamble. These people will display risk seeking behavior.

Indeed, there is a whole range of attitudes an agent could take towards the possibility of getting more than the minimum value of a gamble. And each agent might assess gambles as he does – either weighting improvement over the minimum very highly, or hardly having it count towards the value of a gamble, or anything in between – solely in service of his goal of getting what he values; that is, solely in service of ending up with an outcome he values highly. He might think that taking a gamble with a higher minimum is the most effective way to end up with what he wants, or he might think that taking a gamble with a higher maximum is the most effective way.

Of course, how the possibility of getting more than the minimum is taken into account will depend on the likelihood of this possibility. So, more generally, the value that the possibility of getting more than the minimum adds to a gamble will be the amount which one might get above the minimum, scaled by a *function of the probability* of getting this amount. That is, an agent thinks about a gamble as yielding at least its minimum value, considers the interval by which he might improve over the minimum and the probability of so improving, comes up with a value for this possible improvement, and then adds this value to the minimum value of the gamble.

We can state this formally as follows. Agents who are globally sensitive will have preferences that accord with a *weighted* expected utility function: the desirability of the gamble {A with probability p, B with probability $1 - p$}, where A is at least as good as B, will be u(B) + **r(p)**[u(A) – u(B)], where r is the agent's "risk function" or "weighting function," adhering to the

[27] This describes individuals using a decision rule known as "maximin": take the gamble with the highest minimum. I should note that this rule is usually discussed in the context of decision making under uncertainty, i.e. when an agent does not have a precise probability function, and that most people think of this as an irrational attitude for decision making under risk. The addition of a possibility above the minimum does seem to make a gamble better: for example, it seems that between a gamble that yields $x no matter what and a gamble that yields at least $x but has some chance of yielding a higher amount, we are rationally required to prefer the latter. But notice that it is not clear by how much we are rationally required to prefer it. Indeed, the existence of maximin as a "limiting case" lends support to the claim both that this is how we think about risk and that attitudes in between the limiting cases of maximin and maximax (take the gamble with the highest maximum) might all be rationally permissible.

constraints r(0) = 0, r(1) = 1, r is non-decreasing, and $0 \leq r(p) \leq 1$ for all p.[28] In effect, the interval by which the agent might improve her lot above what she is guaranteed to get is scaled not by her chance of getting the better prize, but by a function of this chance, which reflects her attitude towards risk.[29] Thus the value of a gamble will be the minimum value guaranteed plus the amount by which the agent could do better, weighted by this function of the probability of doing that much better.[30] If this function has a low value, then any possibility of improvement over the minimum will be heavily discounted; thus, the minimum will weigh heavily in the evaluation of the gamble. If it has a high value, then any possibility of improvement over the minimum will be amplified; thus, the maximum will weigh heavily. Since this theory includes a weighting function that reflects the agent's attitude towards risk, I will call it **risk-weighted expected utility theory** (hereafter, **REU**); but there is one more point I need to clarify.

A feature of the standard theory that I have not yet discussed but that is important to preserve is that gambles are not usually stated in terms of the *probabilities* of attaining each outcome, but rather in terms of the states of the world that result in each outcome. According to the tradition following Savage, gambles are functions from states to outcomes: each gamble specifies which outcome that gamble yields in every possible state. So, we can construct a gamble that yields some outcome if a coin lands heads and yields some other outcome if the coin lands tails – but we are not entitled to assume that the gamble generates a 50% probability of each outcome. This is because we are not dealing with objective chance but with *subjective* degrees of

---

[28] Note that I still use "u" and "utility" function in my proposal, even though my 'utility' function is not defined using the axioms that define a standard utility function. I keep the terminology to suggest that my *u* function is trying to capture the same thing as the standard *u* function: the values of various outcomes. If the reader objects, he can call my function *u** and the quantity utility*, but I take "utility*" to be explicating the same concept that expected utility theorists take "utility" to be explicating.

[29] The reader might wonder what exactly the risk function is supposed to 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 it is a desire for more or less risk (if it was the latter, then we would be able to account for it on the standard theory, as I discuss in section five). I admit that I do not have a satisfactory answer to this question. The most informative thing I can say is that the risk function is *not* a propositional attitude, but rather a structural feature of the relationship between beliefs and desires on the on hand and preferences on the other: it explicates an agent's method of aggregating beliefs and desires to arrive at preferences. Standard decision theory itself employs a structural assumption: that this relationship is explicated by an *expectational* function. And if I am on the hook for saying what the relationship is between the structural feature I assume and our folk psychological concepts, then so is the standard theorist. Additionally, however, standard decision theory assumes that this structure is the same across agents. So I have an additional problem: given that the risk function may be different in different agents, I must say what explains this difference. At the moment, I leave this question open to further speculation.

[30] For an example of this, see Appendix A.

belief, and an agent may have any degree of belief in the proposition that the coin lands heads. It is assumed that an agent's degree of belief function, if he can be interpreted as having one, will behave like a probability function, but it is not generally assumed that the decision theorist knows the agent's degrees of belief. Even if the theorist states that the probability of some event is *p*, we cannot assume that the agent has degree of belief *p* in that event (he may think the theorist is lying, or he may think he believes the theorist while his decision making implies otherwise).[31]

Thus, on the standard theory, gambles are of the form {A if event E obtains, B if event ~E obtains}. And the expected utility of the gamble {A if event E obtains, B if event ~E obtains}, where A is at least as desirable as B, u is the agent's utility function, and p is the agent's subjective probability function, is u(B) + p(E)[u(A) – u(B)]. On my theory, then, we calculate the risk-weighted expected utility of that gamble as follows:

> REU({A if E, B if ~E}) = u(B) + r(p(E))[u(A) – u(B)],
>
> where A is at least as desirable as B, and u, p, and r are the agent's subjective utility, probability, and risk functions.

And, of course, most gambles will not have just two possible outcomes, so we need to consider how globally sensitive agents will approach gambles with more than two possible outcomes. The way I've set up the risk-weighted expected utility function in the two-outcome case emphasizes that an agent considers his possible gain above the minimum he is guaranteed (the interval between the low outcome and the high outcome), and discounts that gain by a factor which is a function of the probability of obtaining the gain, a function that depends on how he regards risk. Analogously, in the case in which he will get one of *more than two* possible outcomes, it seems natural that he should again think of the minimum as a baseline; then he should consider the next highest amount he might get above the minimum and his chance of getting *at least* that amount, and determine the value this possibility adds to the gamble; then, treating this value as the "new baseline," consider the next highest amount he might get above this value and his chance of getting at least *that* amount, and determine the value *this* possibility adds; and so forth.

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 receiving at least \$1 more. Furthermore, he has a ¼ probability of receiving at least \$2 beyond that. So the risk-weighted expected utility of the gamble should be u(\$1) + r(½)[u(\$2) – u(\$1)] + r(¼)[u(\$4) – u(\$2)]. This method of calculating gambles is a "bottom up" approach, in

[31] Note that in this paper, I sometimes speak of a gamble that yields an outcome with some specific probability. By this I mean the probability that the *agent* attaches to the event that produces that outcome – his subjective degree of belief in that event.

which the agent thinks of himself as starting off with the worst option, and at each stage takes a gamble to see if he moves up to being guaranteed the next worst option. That is, it emphasizes the minimum utility. We could instead emphasize the maximum utility: treat the agent as if he starts off with the best option and weights the probability of doing no better than the next best option, and so forth. I find the emphasis on the minimum more intuitive. In any case, it makes no formal difference, because the "top down" and the "bottom up" approach will lead to the same weighed-expected utility values, if we transform the r-function accordingly.[32]

In general, the EU of a gamble $\{O_1$ if E_1; O_2 if E_2; … ; O_n if $E_n\}$ will be (with the terms suitably rearranged):

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

And the **risk-weighted expected utility** of a gamble $\{O_1$ if E_1; O_2 if E_2; … ; O_n if $E_n\}$, where O_i is at least as desirable as O_{i-1} (that is, $u(O_1) \leq \ldots \leq u(O_n)$), is:

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

It is worth mentioning that the functional form of my theory is closely related to two existing theories in the economics literature, Schmeidler's Choquet expected utility, a theory of decision making under uncertainty, and Quiggin's anticipated utility, a theory of decision making under risk.[33] Both of these theories weight the utility values of final outcomes by decision weights that are sensitive to the ordering of outcomes. What these decision weights are a function of, though, differs: Choquet expected utility employs a weighting function of states, not of *probabilities* of states; that is, 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 have firm probability judgments (note that if we were to interpret the decision weights as probabilities, then the agent's probability judgments would depend on the gambles he is

[32] Specifically, if r(p) is an agent's risk function when we calculate risk-weighted expected utility using the "bottom up" approach, and if r'(p) = 1 – r'(1 – p) is an agent's risk function when we calculate weighted-expected utility using the "top down" approach, the resulting values will be identical.

[33] Choquet expected utility is described in Hong and Wakker (1996) and Kobberling and Wakker (2003). According to Hong and Wakker, it was originally formulated by Schmeidler in Schmeidler (1989). Anticipated utility was first formulated in Quiggin (1982). Kahneman and Tversky's (1979) prospect theory also employs a weighting function of probabilities, though prospect theory differs in a number of significant ways from all of the other theories I discuss. Note further that I call my theory risk-weighted expected utility theory; but it bears no important relationship to another theory in the literature known as weighted expected utility theory. I do not have room to fully discuss the differences between my theory and other non-expected utility theories here; I merely mention these to situate myself within the literature.

considering).[34] And while anticipated utility attaches decision weights to probabilities, it uses an "objective" probability function – that is, it takes the probabilities as given.[35] My formulation allows that an agent attach *subjective probabilities* to states and then employ 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.[36]

Another difference is conceptual: REU scales *interval differences* in utility by the decision weights, whereas Choquet expected utility and anticipated utility scale utility *values* by interval differences in decision weights. If we rearrange the terms of the equations, we can set r(E) = r(p(E)) to yield Choquet utility, or we can switch to objective probabilities and set r(p) = r(p(E)) to yield anticipated utility.[37] That the equations can be made equivalent may tempt the reader towards thinking that the differences in the functional form (abstracting from the differences mentioned in the previous paragraph about what decision weights attach to) are merely cosmetic. However, weighting the intervals, not the outcomes themselves, highlights that people who are globally sensitive are sensitive to *differences* between utility values of outcomes, rather than engaging in a kind of pessimism (in Quiggin's terminology[38]) by assigning different probabilities to outcomes than their actual probabilities in a manner that depends on how valuable the outcomes are.

My proposal allows that agents evaluate gambles along three dimensions. First, like the standard theory, it allows them to attach subjective values to outcomes: it is up to agents to choose their ends – hence, a subjective utility function. Second, again like the standard theory, it allows them to gauge the relationship between various means and their ends: it allows them to

---

[34] Schmeidler (1989) does include some objective probabilities in a different context from the weighting function, but as the axiomatization in Kobberling and Wakker (2003) makes clear, these are unnecessary.

[35] It also imposes additional constraints: for example, Quiggin assumed – in my terminology – that r(½) = ½ for a gamble with two outcomes. This assumption was crucial for Quiggin, but it seems that it can be easily dropped.

[36] That we can extract p, u, and r from an agent's preferences if they satisfy a particular set of axioms is shown by the representation theorem in my Risk and Rationality (ms).

[37] Peter Wakker in Wakker (1990) notes that Choquet expected utility and anticipated utility are equivalent if we assume an objective probability measure over states (what he calls the "intermediate setup" on p 124) and if we then assume stochastic dominance. Note that my theory could be thought of as a generalization of Quiggin's theory to subjective probabilies. The difference between my theory and Quiggin's is not essential to most of the arguments in this paper (since I do not disagree with standard views about the properties of the subjective probability function), but the difference does entail that the two theories have different axiomatizations; therefore, (1) the justification for each theory via its axiomatization will be different, and (2) my theory will permit the extraction of utilities, decision weights, and degrees of belief from preferences, whereas Quiggin's will only permit extraction of utilities and decision weights. However, since my theory is a generalization of Quiggin's (they are equivalent if probabilities are given; that is, if my theory is restricted to gambles with objective probabilities), everything I say in defense of the rational permissibility of conforming to my theory can be turned into a defense of the rational permissibility of conforming to Quiggin's.

[38] Quiggin (1982), pg. 335.

gauge the likelihood of some particular gamble leading to some particular result – hence, a subjective probability (degree of belief) function. Third, *unlike* the standard theory, it allows them to subjectively judge which means are more *effective* at arriving at their ends. It is up to them to judge which gamble better realizes their goal of getting more money (or their particular ends of getting $50 or, perhaps better by twice, of 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 exactly how these two features of gambles trade off.

Let me briefly say one more thing about why *this* modification of the standard theory, rather than a different modification that takes account of global properties. First, I want to remain within a consequentialist framework – a framework that says that only the outcomes matter – while allowing (unlike the EU theorist) that there are different ways of aggregating possible consequences. Second, I find my proposal to accord with how we think about risk pre-theoretically: how heavily the worst possibility is weighted in one's determination of the value of a gamble seems to be a crude measure of how risk averse one is, and my proposal is a natural refinement of this idea. Finally, though I will not discuss it in detail here, my theory follows from a set of axioms that I find to be reasonable constraints on rational preference – more reasonable, I claim, than those of the standard theory.[39]

On my theory, all of the preferences in the four examples from section one are acceptable.[40] The new equation may look unfamiliar; but, in fact, under certain r-functions, it gives us familiar decision rules. For example, if $r(p) = \{0, p \neq 1; 1, p = 1\}$ for some agent, she uses the maximin decision rule: she pays attention only to the minimum, and any possibility of doing better than the minimum adds no value to the gamble. A similarly simple r-function produces the maximax rule: $r(p) = \{0, p = 0; 1, p \neq 0\}$. If you think that these policies are rationally permissible, then you have some reason to be sympathetic to my proposal, and even if you don't, my proposal suggests a way to argue about them, as well as a way to talk about considerations that might count in against them (e.g. that *r* must be continuous). More importantly, if $r(p) = p$ for some agent, she is a standard expected utility maximizer: she weights an improvement above the minimum by precisely her probability of obtaining it.

---

[39] See my Risk and Rationality (ms). In particular, I accept a relaxation of the sure-thing principle identified by Hong and Wakker (1996) and Kobberling and Wakker (2003).
[40] See Appendix B.

An agent will display risk averse behavior, even though her utility function is linear (or goods are independent), if her r-function is convex.[41] And here we come to the crux of the difference between how the standard theory explains risk averse behavior and how my preferred theory explains it: *on the standard theory, to be risk averse is to have a concave utility function (or, in the general case, a utility function that displays non-independence). On my theory, to be risk averse is to have a convex risk function*. The intuition behind the diminishing marginal utility explanation of risk aversion was that adding money to an outcome is of less value the more that outcome is already worth. The intuition behind my explanation of risk aversion is that adding to the probability of realizing an outcome is of more value the more probability that outcome already has of obtaining. In other words, risk averse people have increasing marginal risk functions: they prefer to make the outcome set less diverse, so to speak. Of course, my theory allows that the utility function is concave (or, indeed, any shape) – but on my theory, this feature, which describes how an agent evaluates outcomes, pulls apart from her attitude towards *risk* itself.

I will not argue extensively for my positive theory here, beyond pointing out that it is intuitive, that it handles the counterexamples, and that familiar theories – including the standard theory – are special cases of it.[42] The main purpose of introducing the theory in this paper is to set up a concrete opponent for the EU theorist. In the remainder of this paper, I will consider arguments on behalf of EU theory: in the next two sections, arguments that agents who violate it are thereby irrational, and in the final section, an argument that it is conceptually impossible for rational agents to violate it.

3 The Sure-thing principle

Since risk-weighted expected utility theory is more permissive than expected utility theory (and, indeed, is a generalization of EU theory), it must lack as least one constraint. Before I present this constraint formally, let me present the intuition it is supposed to formalize.

Suppose I am deciding between two options, say, driving on the highway and driving on the parkway; and I am uncertain about whether it will rain. I might then consult my preferences among the options in the event of rain, and in the event of no rain. Suppose I discover that if it rains, I am indifferent between the two options, since rain will prevent me from enjoying the scenery regardless – on either route, the outcome is "a 2-hour drive without nice scenery." Then I

[41] In particular, an agent with a convex risk function will always undervalue a gamble relative to its expected utility, since $r(p) < p$ for all p. On a related note, an agent will display risk seeking behavior, even though her utility function is linear, if her r-function is concave.

[42] For a more extensive treatment of my positive proposal, see my Risk and Rationality (ms.).

can simplify my decision by only consulting my preferences about what will happen if it does not rain, and letting this preference determine my overall preference. In general, if I have to make a decision, and I don't know what will happen, one way to simplify my decision is to ignore all eventualities that result in the same outcome no matter what I choose. We might call this s*ure-thing reasoning*: if two gambles agree on what happens if one event obtains, then my preference between them must depend only on my preference between what they yield if this event does not obtain.

Savage's "sure-thing principle" formalizes this reasoning as follows:

Sure-Thing Principle (STP) For all X and Y (and for all E): If E and ~E are mutually exclusive and exhaustive events (sets of states), where E is not the null event, then, for all Z and W, I prefer {X if E, Z if ~E} to {Y if E, Z if ~E} iff I prefer {X if E, W if ~E} to {Y if E, W if ~E}.[43]

It is helpful to represent this schematically:

Event	E	~E		Event	E	~E
Deal A	*X*	*Z*		Deal C	*X*	*W*
Deal B	*Y*	*Z*		Deal D	*Y*	*W*

The sure-thing principle says that as long as there is some possibility of *E* happening, then if I prefer deal A to deal B, I must prefer deal C to deal D, and vice versa.[44] The idea is that I have some preference between X and Y, and if I prefer X to Y, then I prefer a gamble that yields X in some states to one that yields Y in those states, as long as the gambles agree on what happens otherwise. This implies that what happens in each state makes a contribution to the overall value of the gamble that is *separable* from what happens in the other states: each outcome makes the same contribution to the value of a gamble regardless of which other outcomes constitute that gamble. Thus the sure-thing principle ensures that preferences among gambles do not depend on irreducibly global properties of gambles, or on the relationships between the outcomes. They do not depend on, e.g., the interval between the best outcome and the worst outcome, the uniformity of the outcomes across states, or the proportion of outcomes reaching a certain threshold. If they did, an outcome's contribution to the value of a gamble would depend on which other outcomes constitute that gamble.[45]

---

[43] Originally due to Savage (1954; 1972). This is the formulation that Susan Hurley (1989) uses, pg. 81.
[44] Assuming preferences are complete, this also implies that if I am indifferent between A and B, then I am indifferent between C and D.
[45] This is because whether an outcome increases the size of the interval between best and worst outcome, or increases the uniformity, or increases the proportion of outcomes reaching a certain threshold depends on what the other outcomes are.

Any risk-weighted expected utility maximizer for whom $r(p) \neq p$ violates the sure-thing principle.[46] Is this principle a constraint on rationality? On the one hand, it does not seem obvious that rationality should rule out caring about irreducibly global properties; but on the other hand, the sure-thing principle seems intuitively compelling.[47] So how should we resolve this dilemma? I submit that the sure-thing principle only seems compelling because we are confusing it with another, more plausible principle.[48]

To evaluate the plausibility of STP, you may have substituted various goods for each variable in the schema, such as:

Event	*Heads*	*Tails*		Event	*Heads*	*Tails*
Deal A	Ice cream	Liver		Deal C	Ice cream	Fruit
Deal B	Pizza	Liver		Deal D	Pizza	Fruit

And you probably thought something like: of course it doesn't matter – when deciding between A and B, or between C and D – what item is in the "tails" column, since if tails comes up, I'll get that no matter what; so my preferences between A and B (and between C and D) should just depend on my preferences between ice cream and pizza. And they should depend on them for the following reason: if I prefer ice cream to pizza, *I'll do at least as well by taking deal A (C) as by taking B (D), no matter how the coin comes up* (and vice versa); and this is true no matter which goods we plug into the tails column. And if I'll do at least as well by taking one deal rather than another and possibly I'll do better, regardless of which state obtains, then I should prefer the former deal to the latter. This last claim is a generally accepted principle:

> **State-wise Dominance: if A is preferred or indifferent to B in all possible states of the world, and preferred in at least one state, then A is preferred to B.**

I have no quarrel with the state-wise dominance principle; and if any principle in the vicinity can rightly be called a restriction on rational preferences, this is it. So if the sure-thing principle

---

[46] We know this is the case because risk-weighted expected utility theory, conjoined with the sure-thing principle, entails that the agent is an expected utility maximizer; that is, that $r(p) = p$. See my Risk and Rationality (ms.).

[47] It is important to note, however, that it doesn't have the same widespread intuitive appeal as, say, transitivity. Patrick Maher mentions that people don't usually change their preferences when told they violate STP, but they do when told they violate transitivity. For the former claim, he cites studies by MacCrimmon (1968), Slovic and Tversky (1974), and MacCrimmon and Larsson (1979); for the latter claim, he cites MacCrimmon (1968). Maher (1993), pg 65, 34.

[48] Of course, there are independent arguments for the sure-thing principle, which I address elsewhere. But one line of thinking suggests that the sure-thing principle needs no independent justification, because it is intuitively compelling (one might say something similar about transitivity). Philosophers who endorse this position include Savage (1972), Harsanyi (1977), and Broome (1991), though Broome does defend STP against counterexamples, as I will discuss in section 5. The point of my remarks here is to respond to this line of thinking and show that one potential intuition in favor of STP actually only favor a weaker principle.

follows from it – or is equivalent to it – then the sure-thing principle is itself a restriction on rational preferences.[49] But STP does not follow from it, as I will now show.

State-wise dominance reasoning cannot be faulted in the example given. However, it fails to apply to all examples in which STP needs to hold, and so STP does not follow from it. In the example, we plugged in (non-risky) outcomes for X, Y, Z, and W; but for STP, these variables actually range not only over outcomes, but also over *gambles*. And when gambles are plugged in for the variables in the STP schema, there is not always one option that is better in every state, so we cannot always apply dominance reasoning.[50] For example, consider the following schematic diagram of the choice the Allais agent must make, where the prize is determined by drawing one out of a hundred lottery tickets:

Ticket	1	2-11	12-100		Ticket	1	2-11	12-100
L_1	*$0*	*$5m*	*$0*		L_3	*$0*	*$5m*	*$1m*
L_2	*$1m*	*$1m*	*$0*		L_4	*$1m*	*$1m*	*$1m*

Take *E* to be the event in which ticket 1-11 is drawn. L_1 and L_2 agree on what happens in ~*E*, as do L_3 and L_4. Therefore, in order for STP to be satisfied, an agent's preference between deal L_1 and deal L_2 must be the same as that between deal L_3 and deal L_4: that is, the common Allais preferences violate STP. Importantly, though, for neither of the choices – the choice between L_1 and L_2 or the choice between L_3 and L_4 – is it the case that I can do better *no matter what* by taking one gamble rather than the other: if ticket 1 is drawn, L_2 and L_4 are better, and if one of tickets 2-11 is drawn, L_1 and L_3 are better.

The sure-thing principle is much stronger than a principle that says "take A over B if you can do better no matter what by taking A rather than B." Indeed, what the sure-thing principle requires is that if I **all-things-considered** prefer X to Y, I must prefer that X rather than Y be a sub-gamble of any gamble I take. But one thing that many people consider when deciding which gambles to take is their global properties. And X and Y might help instantiate different global properties, depending on which gamble they are embedded in. In this example, let X be a 10/11 chance of receiving $5m, determined by drawing one of tickets 1-11, and let Y be receiving $1m

[49] Indeed, this is the reasoning that John Harsanyi gives in support of the sure-thing principle in response to Watkins's (1977) criticisms of expected utility theory: "[The sure-thing principle] is essentially a restatement, in lottery-ticket language, of the *dominance principle*…The dominance principle says, If one strategy yields a better outcome than another does under *some* conditions, and never yields a worse outcome under *any* conditions, then always choose the first strategy, in preference over the second. On the other hand, the sure-thing principle essentially says, If one lottery ticket yields a better outcome under *some* conditions than another does, and never yields a worse outcome under *any* conditions, then always choose the first lottery ticket. Surely, the two principles express the very same rationality criterion!" Harsanyi (1977), pg. 384.

[50] Edward McClennen makes this same point in McClennen (1983), pg. 176-8.

for certain. An agent might all-things-considered prefer Y to X because although X has some positive features (e.g., \$5m if one of tickets 2-11 is drawn), Y's positive features (e.g., \$1m no matter what) are more compelling. However, when X and Y are embedded in gambles which yield \$0 if one of tickets 12-100 is drawn (i.e., in gambles L_1 and L_2), the positive features of the gamble in which X is embedded might now outweigh the positive features of the gamble in which Y is embedded (features which no longer include receiving a guaranteed \$1m). STP is only intuitively compelling if global properties are not part of what agents should consider when making decisions. Therefore, supporting STP via intuition straightforwardly begs the question against caring about global properties.

The reason, I suspect, that STP initially sounds so intuitive is because the applications we immediately call to mind are those in which the variables range over outcomes. However, in restricting our attention to these applications, we are actually considering a much weaker principle, SWD. STP does not follow from SWD, because there are some applications of STP in which neither option is such that the agent will do at least as well in every state by taking that option as by taking the other. Interestingly enough, risk-weighted expected utility maximizers obey a principle intermediate between the two, which seems to capture both the idea that agents should use sure-thing reasoning when gambles do not differ in important global properties and the idea that agents can care about global properties of gambles. That is, risk-weighted expected utility maximizers obey a restricted version of STP.[51]

What is at issue is whether rational agents are required to obey STP even when gambles differ in global properties. Therefore, to see whether STP should be endorsed in full, we need to consider the full range of cases. Since the principle is not justified via intuition in all of these cases, the EU theorist needs to secure STP via argument. Various arguments for STP can be found in the literature; I do not think they succeed, but I will not take up that discussion here.[52] At the very least, I hope I have shown that STP is not an obvious principle of rationality like, say, transitivity.

4 Should we care about expectation?

---

[51] In particular, they obey what Hong and Wakker (1996) call the Comonotonic Sure-Thing Principle. This principle is the restriction of the sure-thing principle to gambles within the same comoncone; that is, gambles such that the order in which the agent ranks the events (in terms of which prizes they yield) is the same – gambles that order the events in the same way as each other. See my Risk and Rationality (ms).

[52] I discuss three important arguments for STP, the argument from the non-aversiveness of knowledge and the arguments from consistency over time and from consistency at a time, in my Risk Aversion and Rationality (ms). Helpful presentations of these arguments appear in Machina (1991), Maher (1993), and McClennen (1983).

Let me turn to an argument that rational agents must obey (non-risk-weighted) expected utility theory as a whole; of course, this theory entails the sure-thing principle, so if this defense of expected utility theory succeeds, then as a consequence, rational agents must obey the sure-thing principle.

This defense tries to make explicit the connection between utility and *expected* utility. As mentioned in section 1, under the formalistic interpretation, utility is not a real quantity. Rather, it measures the strength of our preference: whatever the agent's utility function (assuming a utility function can be defined), he prefers an outcome that has higher utility to one that has lower utility. In this section, I will explore an argument that, given that an agent prefers the outcome with the highest utility, an agent must prefer the *option* with the highest *expected* utility. I call this claim the Expecting claim:

> **(EX) If option A has higher *expected* utility than option B for a rational agent, then that agent prefers option A to option B.**

This claim is neutral between the realistic and formalistic interpretation: if utility cannot be defined except via some theory that links preferences with utility, then we can read the Expecting claim as stating that rational agents must conform their preferences to a theory that identifies preference with expected utility (rather that, e.g., a theory that identified preference with risk-weighted expected utility). In other words, we can read it as stating that the theory from which utility is derived must have it that we are indifferent between a basic outcome and a gamble whose average utility is equal to the utility of that basic outcome.[53] The Expecting claim links *expected* utility and *actual* utility, regardless of whether this link is used to necessitate that an agent make certain choices or to constrain how a correct theory defines his utilities.

The Expecting claim is historically significant: it can be found in the earliest formulations of decision theory by Blaise Pascal, and both Maurice Allais and Hilary Putnam take it to be the background to their arguments.[54] Pascal's original innovation in decision theory was to propose that the *monetary value* of a stake in a game should be equal to its expected value.[55] Joyce notes two attractive features of Pascal's proposal that if a game ends prematurely, each gambler should be given as his share of the pot the monetary expectation of his stake. The first is that a player's

---

[53] Perhaps this argument seems less compelling on the formalistic conception of utility. However, this argument needs to be addressed because (1) I take it that on the formalistic picture, utility is still meant to measure value in some sense, so we can make sense of the idea that value should coincide with average value, and (2) a common criticism that arises is that non-expected utility maximizers will do worse over the long run than expected utility maximizers, so this section can be taken as a response to that criticism.

[54] See Joyce (1999); Allais (1953); Putnam (1986).

[55] Interestingly enough, the move from talking about expected value to expected *utility* came in response to noted widespread risk aversion in money.

share of the pot depends only on two (objective) features of his situation: the amounts of money he might have won and the probabilities he had of winning those amounts. The second is that the amounts of money that each of the players gets add up to exactly the amount of money in the pot. Neither of these features entails that the players should be indifferent between receiving their share of this division and continuing the game, although it may be difficult to argue that the division favors any particular player. More importantly, though, neither of these features is preserved when we move from equating monetary value with expected monetary value to equating *utility* with expected utility:[56] whereas the expected monetary value of a player's stake depends only on objective amounts of money and probabilities, the expected utility of his stake also depends on the (subjective) utility he assigns to each amount of money; and whereas the expected monetary values of each stake always add up to the total amount of money in the pot, the expected utilities of each stake clearly do not sum to the utility of the money in the pot.[57]

So why do we have reason to prefer the gamble with the higher expected utility? Clearly, it cannot be because we are likely to do better in any particular instance by choosing a gamble with higher expected utility than we would by choosing a gamble with lower expected utility; we are not. For example, if we are choosing between a gamble that yields a prize worth 200 utiles if ticket 1 is drawn, and nothing if tickets 2-100 are drawn, and a gamble that yields a prize worth 1 utile if tickets 1-100 are drawn, then the former has a higher expected utility than the latter, even though I am *likely* to do better by taking the second gamble; that is, in 99 of the 100 (equally likely) possible states of the world, I will do better by taking the second gamble than by taking the first. Higher expected utility does not entail a likelihood of higher utility, because expected utility is not just responsive to the possibility of doing better: the size of the amount by which I stand to do better can outweigh the low probability of obtaining that amount.

Since expected utility is just an average of the utility values of the gamble in all possible states of the world, the reason to care about it must make reference to more than one result – e.g. to what will happen over the long run or to what will happen in all possible worlds. I set this latter possibility aside[58] and consider the claim that rational agents should make decisions that maximize expected utility because *over the long run, an expected utility maximizer will end up*

---

[56] Joyce (1999), pg. 9-11.

[57] Unless utility is linear in money and equivalent for every player. There is a problem about how to compare utility values across people: since utility is fixed only up to scale and unit for each individual, it is difficult to find a criteria for selecting a member of the class of utility functions that represent me and a member of the class of utility functions that represent you that have the "same" scale and unit. Therefore it does not even make sense to "sum" the expected (subjective) utilities of each player's stake.

[58] In my Risk and Rationality (ms), I consider (and reject) the "possible worlds" defense of EU theory: that rational agents should make decisions that maximize expected utility because this amounts to making decisions that maximize the total utility across possible worlds.

with higher utility than someone following another strategy, e.g., than someone who uses maximin.[59] More precisely, assume that an agent faces a series of choices identical to the original decision problem, and that he must choose the exact same gamble each time. Then, as the number of (identical) choice situations becomes larger and larger, it grows more and more probable that taking the gamble with the highest expected utility (each time) will lead to a higher amount of average utility than taking any other gamble (each time); this is true by the law of large numbers.[60] If Ralph were to participate in identical instances of the referee's gamble ad infinitum, then over the long run, he'd (with near certainty) get the exact same average number of gloves and Elvis stamps by taking deal 1 as by taking deal 2 – either way, he'd average half an Elvis stamp and half a pair of gloves per instance of the gamble.[61] So he shouldn't prefer one deal to the other. Similarly for Margaret: over the long run, she'd average 50 utiles on each instance of the gamble whether she took \$50 or the coin flip between \$0 and \$100; so she shouldn't prefer the former to the latter.[62] And for the Allais agent and Watkins: depending on how they values the monetary outcomes, either L_1 and L_3 will lead to a higher average utility over the long run than L_2 and L_4, respectively, or vice versa; and at least one of the rules R1 or R2 will recommend choices that will lead to a lower average utility over the long run.

So, the argument goes, if one expects to participate in a gamble repeated ad infinitum, then maximizing expected utility will lead (with near certainty) to having a higher utility on average from each instance of the gamble (and so to a higher total utility). In other words, it will lead to doing better for oneself. I will first respond to this argument by showing that it does not prove enough: it does not show that one should maximize expected utility in any particular instance. I will then argue that if this fact proves anything, it proves too much: if we can justify expected utility maximization by reference to what will happen in the long run, then we can also justify other decision-making strategies, such as maximin, in this same way; and furthermore, we can even show that agents should follow the original expected *value* theory instead of expected utility theory.

---

[59] I use maximin as an example because it is a particularly clean case and because it is a "limit case" of my proposal.

[60] The law of large numbers states that as the number of trials of a random variable approaches infinity, the probability that the average of the trials is more than ε from the mean value of the variable approaches zero. By average utility, I mean the sum of the utility values that result from each gamble divided by the total number of gambles.

[61] More precisely, the value of his average prize would be the same either way: half the value of the gloves plus half the value of the Elvis stamp.

[62] Taking for granted that Margaret values money linearly, and setting the unit and scale (harmlessly) so that \$50 is worth 50 utiles, \$0 is worth 0 utiles, and \$100 is worth 100 utiles.

For my first response, I will grant to the EU theorist that *near* certainty of doing better is enough to entail that if a situation is to be repeated ad infinitum, then one should adopt the policy of maximizing expected utility. Still, it is not obvious that one should care about maximizing expected utility in situations that are not to be repeated ad infinitum. What will happen in a very long amount of time need not resemble what will happen in a much shorter time.

Maurice Allais, who himself rejects the expectational theory in favor of a theory that takes into account the spread of possible outcomes resulting from an action, argues that hypothetical value over the long run is not a good measure of actual value to the agent.[63] He makes two points that are relevant here: first, no agent can actually participate in an unlimited number of gambles with an unlimited bankroll. Second, many or most decisions will be isolated events, either unrepeatable by nature or unlikely to be repeated.[64] Hilary Putnam also makes this second criticism in his "Rationality in Decision Theory and Ethics."[65] He asks us to consider whether someone who does not have much life left to live should be swayed by considerations of what would happen in the long run, and concludes that "If my *only* reason for believing that I should be reasonable were my beliefs about what will happen *in the long run* if I act or behave reasonably, then I would have absolutely *no* reason…to think it better to be reasonable in an unrepeatable single case."[66] Putnam still thinks that maximizing expected utility is the only rational thing to do, even if the situation is unrepeatable; however, he thinks that the only response available to the EU theorist is to say that he cannot and need not justify his theory. Though this response might satisfy a die-hard believer in EU theory, it won't convince the agnostic to accept it.

The EU theorist might try to respond to both criticisms by pointing out that even though an agent will not face the exact same decision over and over again, he will most likely face an extraordinary number of decisions throughout his lifetime, and adopting a policy of maximizing expected utility will, with increasing likelihood, get him the most utility on average. However, this is not what the law of large numbers says: it only talks about the expected value of a single variable. Moreover, refusing to maximize expected utility in a single instance will "wash out" in the long run – it will only change the speed of convergence to the average – so there is no reason to be always bound by the policy.[67]

---

[63] Allais (1953), pp. 116-118.
[64] A particularly clear case of a gamble that is unrepeatable by nature is Pascal's famous wager.
[65] Putnam (1986), pp. 3-16.
[66] Putnam (1986), pg. 9, 12. Italics Putnam's.
[67] Granted, this problem comes close to problems like the "tragedy of the commons," in which it is the case that any particular instance of deviation will wash out, but widespread deviation will not. Perhaps it is too much to require that the EU theorist solve this problem in order to justify his theory using this defense.

It seems this line of reasoning will not work for the defender of EU theory. But let us grant, for the sake of argument, that we can justify a single instance of a decision-making strategy (such as expected utility maximization) by showing that if the situation were to be repeated, it would be preferable over the long run to adopt that strategy each time. Then, I argue, we can prove too much: for if this argument (from what would happen in the long run to what an agent should do in some particular instance) is valid, then advocates of other decision rules can use a parallel argument to justify *those* rules.

I've already pointed out that because of the way utility is defined, everyone has a preference for the outcomes that have higher utility over those that have lower utility. And the expected utility maximizer interprets taking the most effective means to satisfying this preference to require that the agent choose so that, with near certainty, the gambles he takes average out to the highest utility. However, there are other standards of success. The maximinimizer can interpret taking the most effective means to require that the agent get the most he can be *guaranteed* to end up with – and, in the long run, choosing the gambles that a maximin strategy recommends will give him as high a guarantee (and usually higher) as choosing any other gambles. Unless the expected utility maximizer literally takes infinity instances of a gamble, there is still some (vanishingly small) chance that things will turn out as badly as they could turn out in every instance. After any finite number of trials, an agent might end up with the minimum value of his chosen deal in every trial – the coins may land TH each time, and Ralph may end up with nothing – and the minimum value of the maximinimizer's chosen deal will be at least as high as, and sometimes higher than, that of the expected utility maximizer's chosen deal. In other words, the expected utility maximizer is succeeding by his own interpretation of the standards, and the maximinimizer is succeeding by *his*.

The second point I want to make towards showing that this argument proves too much is that even if we accept the EU maximizer's interpretation of the standards, we can produce an argument of the same form but for an incompatible conclusion. This argument shows that an agent who cares about getting the highest average of the quantity he prefers (were the gamble repeated ad infinitum) should not be maximizing expected utility at all – rather, he should be maximizing expected *monetary value*. The argument runs as follows: assume an agent cares about doing better on average, and that he prefers larger amounts of money to smaller. If the gamble is to be repeated ad infinitum, then, again by the law of large numbers, it will grow increasingly likely that taking the gamble with the highest expected *monetary value* leads to a larger average amount of *money* than taking any other gamble. And since the agent prefers larger

amounts of money to smaller (indeed, larger amounts of money have higher utility), he should maximize expected value.

The original argument and my argument of the same form have contradictory conclusions. It seems we have shown that agents interested in getting higher utility in repeated gambles over the long run should both maximize expected utility and maximize expected monetary value. But these do not come to the same thing (unless utility is linear in monetary value). To spell out this contradiction more precisely, the law of large numbers implies that taking the gamble with the highest expected utility leads to the highest average utility (per trial), and that taking the gamble with the highest expected monetary value leads to the largest average amount of money (per trial).[68] And since the average amount of quantity *x* in *n* trials is just the total amount of *x* divided by *n*, then after *n* trials, taking the gamble with the highest expected utility should lead to the highest total utility, and taking the gamble with the highest expected monetary value should lead to the largest amount of money. But larger amounts of money have more utility; so we have an argument that two incompatible strategies each lead to higher total utility than any other strategy (including each other).

What is going on here? As it turns out, a gamble can yield the highest average utility on each trial without having the highest total utility – but only if by average we mean we *first* take the utility of the monetary prize a gamble yields on *each* trial (ignoring which other prizes the agent has won on previous trials), and *then* sum these utilities (then divide by the total number of trials), and by total we mean *first* add the monetary prizes together and *then* take the utility of that number. That a repeated gamble yields a higher average utility on each trial, where the utility of a gamble on a particular trial is calculated without regard to the results of the other trials, does not entail that someone taking this repeated gamble will be left with a better collection of prizes at the end. This is because, as already stressed, the utility of two prizes together is not always the sum of their utilities (e.g., u(\$50) + u(\$50) ≠ u(\$100)); prizes are not *always* independent. Utility values of prizes, unlike monetary values, depend on how one does in other gambles (especially if utilities diminish marginally).

To make this point concrete: let us assume an agent's utility function assigns a utility of 0 to \$0, 0.6 to \$25, and 1 to \$100. And assume he is offered a choice between option A (\$25), which has an expected utility of 0.6, and option B, a fair coin flip between \$0 and \$100, which has expected utility 0.5. Option A had a higher expected utility, and, after, say, 1000 trials, it will yield an average utility of 0.6, a total monetary amount of \$25,000, and a total utility of

[68] I will omit the caveat that this is only *nearly* certain (i.e. only with increasing likelihood), because it makes no difference to what follows.

u($25,000). Option B has a higher expected *value*, and after 1000 trials, it will yield an average utility of 0.5, an average monetary value of (roughly) $50, a total monetary amount of (roughly) $50,000, and a total utility of u($50,000) – a higher total utility than Option A.

Does this show that EU maximizers do worse over the long run than expected value maximizers? No, because this argument isn't in fact faithful to how an EU maximizer values repeated gambles. When faced with a choice among gambles of which one gamble is to be repeated ad infinitum, the expected utility maximizer will calculate the expected utility of each collection he is considering (that is, each collection that consists of repeated instances of one gamble). As emphasized above, since the utility of two prizes together is not always the sum of their individual utility values, the expected utility of this collection will not be the number of gambles multiplied by the expected utility of the individual (non-repeated) gamble. The gamble he will choose to be repeated is the gamble corresponding to the *collection* that has the highest expected utility. If the gamble is to be repeated ad infinitum, this will be the collection corresponding to the gamble with highest expected monetary value (in our example, option B). Incidentally, the REU maximizer will approach a repeated gamble in a similar manner: he will calculate the REU of each collection and choose the gamble corresponding to the collection with the highest REU; and if the gamble is to be repeated ad infinitum, this, too, will be the collection corresponding to the gamble with the highest expected monetary value (again, option B). So both the EU maximizer and the REU maximizer will choose option B repeated ad infinitum rather than option A repeated ad infinitum (provided they prefer more money rather than less), even though they might both choose option A if it is not to be repeated. The value of repeating a gamble will not be the number of gambles multiplied by the value of the individual gamble for either agent: for the EU maximizer because values are not always independent (utility is not always additive), and for the REU maximizer because the probability of a final outcome depends on all the gambles in play (relatedly, the r-function is not always additive).

Indeed, if one can anticipate a gamble repeating with no limit, maximizing expected value, maximizing expected utility, and maximizing risk-weighted expected utility all recommend the same strategy for agents who prefer more money rather than less. My counterargument does not show that the expected utility maximizer will behave irrationally by his own lights. What it does show is that the initial argument is *unsound*: maximizing expected utility in a single instance cannot be justified by citing the purported fact that if the gamble were

to be repeated and he were to make the same choice every time, the expected utility maximizer would do better (by his own standards) than a decision maker following any other strategy.[69]

This argument was doomed from the beginning. Discussing what an agent should do in a repeated gamble in order to show that an agent should not care about risk straightforwardly begs the question. Repeating a gamble is a way to minimize risk, because repeating a gamble reduces the variance; in the limit, the variance goes to zero. Thus, it is obvious that someone who cares about properties like variance will not behave in the non-repeated situation as he will in the repeated situation. Justifying a single instance of expected utility maximization by reference to what happens in the long run obscures the very thing at issue: namely, whether properties of a gamble other than its mean value can matter to a rational agent.

5 Individuation of Outcomes

In the first section, I gave several examples in which people tend to violate expected utility theory. Of course, people's preferences only violate EU theory if we assume that the outcomes are as stated in the examples. In this section, I will address a common move that expected utility theorists make to salvage their theory in light of examples of preferences that seem to violate it. This move consists in individuating outcomes so that the preferences in question do obey the axioms of the theory.

To justify this move, the theorist points out that the outcomes in the examples are not stated in enough detail to capture everything of value to the agent. In my examples, the theorist might claim that it is relevant to the value of the outcomes that things could have turned out otherwise. Specifically, it is relevant to the value of not receiving any prize that Ralph might have gotten something had he chosen differently, because he would feel regret, which itself has negative value. Thus, the correct description of the outcomes is really more fine-grained than as initially presented in the problem; the options are:

	HH	HT	TH	TT
Deal 1	*Elvis stamp*	*Elvis stamp and gloves*	***Regret***	*Gloves*
Deal 2	*Elvis stamp*	*Elvis stamp*	*Gloves*	*Gloves*

[69] Is there another way to make an argument from the long run, e.g., by considering the gamble to be repeated as one with set *utility* values? That is, consider a gamble whose prizes depend on what the agent won in the previous round, so that they represent the same utility change in his overall wealth, and then argue that an agent will do better by taking the gambles that (each) maximize expected utility rather than any other gambles. Setting aside the issue that the gambles offered in each round depend on one's winnings in the previous round (and so which gambles one is offered depends on which choices one makes), again the REU maximizer will choose the *collection* with the highest REU, which in the limit will be the collection corresponding to the gamble with the highest EU.

The outcome of deal 1 in the TH state is not simply the status quo, so Ralph's preference for deal 2 over deal 1 no longer violates expected utility theory.[70]

Alternatively, if this does not seem to correctly describe Ralph's preferences, I've said that Ralph prefers deal 2 to deal 1 because he is sure to win a prize. The EU theorist could say that this is because the fact that he will get a prize for certain adds value to each outcome in deal 2, perhaps because Ralph enjoys anticipating a prize: e.g. "gloves and anticipating some prize" is better than "gloves." We could describe the outcomes to take this into account:

	HH	HT	TH	TT
Deal 1	*Elvis stamp*	*Elvis stamp and gloves*	*Nothing*	*Gloves*
Deal 2	*Elvis stamp* ***& anticipation***	*Elvis stamp* ***& anticipation***	*Gloves* ***& anticipation***	*Gloves* ***& anticipation***

Again, Ralph's preference for deal 2 over deal 1 no longer violates expected utility theory.[71]

The same strategy works in response to the Allais paradox, and indeed is standardly employed. The theorist points out that it is relevant to the value of receiving \$0 in L_3 that you might have received \$1m had you chosen differently, because you feel regret, which itself has negative value. Thus, the correct description of deal L_3 is "\$1,000,000 with probability 0.89, \$5,000,000 with probability 0.1, \$0 *and regret* otherwise." If these are the outcomes in L_3, then the common preferences no longer violate expected utility theory.[72] Again, there are other ways to individuate outcomes to salvage the theory: L_4 might be "\$1m with probability 1, without the worry of gambling." I do not mean to privilege any particular description; the point is that if some of the states in which the agent receives \$1m (or \$0) are different from some of the other states in which she receives \$1m (\$0), then the classic preferences no longer violate the sure-thing principle, or indeed expected utility theory.[73] There are other classic examples of re-individuation

---

[70] EU(deal 1) = ¼u(stamp and gloves) + ¼u(stamp) + ¼u(gloves) + ¼u(regret), and since the term "u(regret)" does not appear in the equation for the value of u(deal 2), there is no necessary connection between the two.

[71] EU(deal 2) = ½u(stamp and anticipation) + ½u(gloves and anticipation); again, neither of the terms (nor "u(anticipation)" alone) appear in the equation for the value of u(deal 1).

[72] Expected utility theory can now capture the preference for L_1 over L_2 and L_4 over L_3: there is no contradiction between the following two equations:
0.1(u(\$5m)) + 0.9(u(\$0)) > 0.11(u(\$1m)) + 0.89(u(\$0))
u(\$1m) > 0.89(u(\$1m)) + 0.1(u(\$5m)) + 0.01(u(\$0 and regret))
On the contrary, it is easy to find utility functions that satisfy both equations.

[73] This strategy can also be employed for Margaret's preferences, although the re-individuation may be more complex: instead of a coin flip between \$0 and \$100, the gamble is a coin flip between \$0 with regret and \$100 – or "\$0 as the result of a gamble that had a 50% probability of \$100" and "\$100 as the result of a gamble that had a 50% probably of \$0." And similarly for Watkins: instead of a coin flip between, say, \$10 and \$40, the outcomes of the gambles he faces are things like "\$10 as the result of a gamble that had a 50% probability of \$10 and a 50% probability of \$40" or "\$10 as the result of a gamble that had a 50% chance of one sum of money and a 50% chance of another which is \$30 larger." Formulations of the latter type may be necessary if an agent is globally sensitive and we know a lot of her preferences, e.g. if there is some

of outcomes in response to purported violations of the standard axioms, but I will not discuss them here.[74]

There are two motivations for re-individuation in decision theory. The first is the view that the outcomes are genuinely under-described in many of these problems: the initial descriptions ignore features of the problems that we think are important, and we cannot normatively assess an agent without taking these features into account. The second comes from a philosophical picture that sees the primary advance of decision theory to be its providing us with a way to discover an agent's beliefs and desires: given the connection between preferences on the one hand and beliefs and the desires on the other – a connection that representation theorems make explicit – we can determine an agent's beliefs and desires through knowing only his preferences.[75] Since the aim of decision theory, on this picture, is to interpret how an agent sees the world, we may also interpret how he carves up the outcomes. However, on both this picture and on the picture in which decision theory is a tool for normatively assessing agents, there must still be some constraints on how outcomes are individuated. Otherwise, when does it all end? If we can always re-describe outcomes whenever an agent's preferences seemingly conflict with decision theory, then the theory will not tell us anything substantive about which preferences are rational, or about how we should interpret agents.

For example, assume we have an agent who chooses chicken over steak at one time, and steak over chicken at another. One way to interpret her might be to say that she prefers steak to chicken when the tide is high in Alaska, and chicken to steak when it is not. This interpretation makes her consistent, but if she does not have any knowledge of the Alaskan tide, and if this event is wholly unconnected with her decision, then this interpretation will not help us get at her real beliefs and desires. If we are interested in decision theory as a framework for discovering an agent's beliefs and desires, we want to rule out certain interpretations of her beliefs and desires, interpretations that don't make sense of what she is doing. And if we are normative theorists, we also want to rule out some sets of preferences, preferences that the agent seems to have no reason to have. Similarly for the agent who looks to decision theory as a practical guide to her own actions: upon realizing that she prefers chicken to steak when the tide is high, she need not be in doubt about her preference when it is low. In other words, we might think that "chicken when the

---

utility function and r-function (that is not the identity function) that represent her preferences under risk-weighted expected utility theory.

[74] See, e.g. Pettit (2002). Pettit also cites Peter Diamond (1967). I discuss differences between these cases at length in "Risk without Regret" (unpublished paper).

[75] Ramsey originally formulated his decision theory in response to problem of measuring degrees of belief (see Ramsey (1926). See also Lewis (1974) and Hurley (1989), especially chapters 4 and 5. Here, Hurley cites, among others, Lewis (1983) and Davidson (1986).

Alaskan tide is high" and "chicken when the Alaskan tide is low" should not count as different outcomes for rational agents. It is clear that we need a restriction on when outcomes can be finely individuated and when they cannot be.

Broome has such a restriction.[76] He calls it the Principle of Individuation by Justifiers:[77]

> **(PIJ) Outcomes should be distinguished as different if and only if they differ in a way that makes it rational to have a preference between them.**

So "chicken when the Alaskan tide is high" and "chicken when the Alaskan tide is low" are the same outcome, since it is not rational to have a preference between them.[78] Therefore, they must be interchangeable – if one is preferred to steak, the other must also be – and the agent must be indifferent between them. Actually, strictly speaking, Broome thinks that *any* two outcomes can be individuated, but that rationality requires an agent to be indifferent between some of them. But as he points out, (PIJ) implies a rational requirement of indifference: unless two outcomes differ in a way that makes it rational to have a preference between them, an agent must be indifferent between them.[79] Unless chicken-at-high-tide and chicken-at-low-tide differ in a way that makes it rational to prefer one to the other, the agent must be indifferent between them, and therefore, again, they must be interchangeable. Since the difference between the two principles does not matter for our discussion, we will follow Broome in using (PIJ) rather than an indifference principle, which he does for expository reasons.

With this principle in place, Broome has an ingenious argument that no rational agent can violate the sure-thing principle. Remember, we only need decision theory to accommodate Ralph's preferences if they are rational. Broome's argument, if it succeeds, sidesteps the debate about the correct theory of decision-theoretic rationality, since he argues that agents who are rational, *however we spell out this concept*, cannot violate STP. We will see how this argument works in the case of the Allais paradox, and then generalize it. Broome writes:

> "All the [rationalizations of the Allais preferences] work in the same way. They make a distinction between outcomes that are given the same label in [the initial presentation of the options], and treat them as different outcomes that it is rational to have a preference between. And what is the argument that Allais's preferences are inconsistent with the sure-thing principle? It is that all the outcomes given the same label [initially] are in fact the same outcome. If they are not…[the decision problem] will have nothing to do with the sure-thing principle. Plainly, therefore, the case against the sure-thing principle is absurd. It depends on making a distinction on the one hand and denying it on the other."(Broome 107)

---

[76] Philip Pettit (2002) has a slightly different restriction. I discuss this in "Risk without Regret" (unpublished paper).
[77] Broome (1991), pp. 103.
[78] It might sometimes be rational to have a preference between them, in which case they would count as different outcomes. The point is that whenever we want to say that two outcomes are not (relevantly) different, we can point out that it is not rational to have a preference between them.
[79] Broome (1991), pp. 103-104.

Broome points out that in order to use the Allais paradox to show that rational agents violate the sure-thing principle, one needs to show both that the common Allais preferences are rational and that they violate the sure-thing principle. In order to show that they violate the sure-thing principle, one must show that the outcomes that appear the same in the original choice problem should not be individuated (i.e., that the original choice problem really is an instance of the sure-thing schema). That is, we need to show that there is *no rational difference between the outcomes*. However, if the preferences are rational, it must be true that there is a difference between some of the outcomes that appear the same – a difference that the agent can rationally care about.

I mentioned above that there are several different ways in which the decision theorist can individuate outcomes in the Allais paradox so that the common preferences do not violate standard decision theory. And, Broome presumes, any way of rationalizing the preference for L_1 over L_2 and L_4 over L_3 will make reference to a difference in some of the outcomes: it will be interpretable as one of the ways to individuate outcomes more finely than they are individuated in the original set-up of the problem. If it is rational for the agent to distinguish between those outcomes, then the decision theorist can also distinguish between them, and the agent's preferences will be rational but will not violate STP. On the other hand, if it is not rational for the agent to differentiate those outcomes, then the decision theorist cannot do so either, and the agent's preferences will violate STP but will not be rational; so it will not matter that they violate the theory.

This argument can be extended to any purported violation of the sure-thing principle. If the agent can justify a purported violation of STP, then (by Broome's line of reasoning) it will be by reference to differences among some of the outcomes that initially appear the same; but then the decision theorist will point out that the preferences over options with the newly described outcomes do not actually violate the sure-thing principle.

Broome's argument is clever. However, it makes an assumption that (I argue) is false: *that any way of arguing that the common Allais preferences (or any STP-violating preferences) are rational relies on making a distinction between outcomes that are initially given the same label*. If this is not the case, then Broome's argument does not go through: for if there are decision situations in which it is rational to have preferences that violate the sure-thing principle, but in which it is not rational to have a preference between outcomes that are, as stated, the same – and therefore in which the decision theorist must not individuate preferences any more finely

then they are already individuated – then it is rational to violate the sure-thing principle, and individuation will not save standard decision theory.[80]

Broome's assumption, I claim, is not true in the case of the Allais preferences: it is not that people have rational preferences between \$1m as the result of a gamble and \$1m without the worry of gambling, or between \$0 with regret and \$0 without regret, but rather that \$1m and \$0 contribute something different to gambles L_3 and L_4 than they do to L_1 and L_2. *My justification for the Allais preferences does not depend on distinguishing between outcomes that appear identical, but on distinguishing how identical outcomes contribute to different gambles.* I think that for many agents, once the gamble has been decided and the agent is left holding \$0, \$1m, or \$5m, what the agent could have gotten will not make any difference to the value of her actual winnings.[81] And similarly with Ralph: once we know whether he has an Elvis stamp, gloves, both, or only what he had before he took the gamble, there may be nothing more to know about the value of his holdings. But when Ralph and the Allais agent approach the gambles before the results of the gambles are known, how outcomes are distributed over the various states of nature makes a difference. Identical outcomes may not affect the value of two gambles equally, since the part they play in a gamble – the global properties they help instantiate – depends on other possible outcomes of the gamble.

Broome's argument fails because he assumes that any differences in what outcomes contribute to a gamble must be differences in the outcomes themselves. He makes this explicit: "The value that Allais associates with interactions between states, is really *dispersed* amongst the states themselves. In this, I was faithfully following all the available rationalizations of Allais's preferences; they all depend on feelings of some sort…Nearly all the published counterexamples to the sure-thing principle are like this."[82] Surely there is a way of reading any counterexample to STP as having different outcomes than originally thought; but that is not always the correct way to read them. The properties that globally sensitive agents value – like low variance or a high minimum – are properties that attach to gambles before their results have been determined. After such an agent wins his prize, he may not care whether it was the result of a risky gamble or was simply given to him. This is why the values of these global properties cannot be dispersed among

---

[80] At least, it won't save it if we employ a principle like (PIJ). It could still save a decision theory that is not supposed to have any normative content (e.g. one whose axioms are trivially true and in which re-individuation is always allowed).

[81] Or at least will not make a large enough difference to the value of her winnings to explain the preference shift: in order for regret to explain the Allais preferences, it has to have not just a negative value to the agent, but a large enough negative value to balance out the value difference between a 10/11 chance of \$5m and a sure-thing \$1m. Thanks to Dan Greco for making this latter point.

[82] Broome (1991), pg. 110.

the states; they will not truly be the values of the outcomes by themselves.[83] Since riskiness is a property of a gamble before its result is known, it need not, so to speak, leave a trace in any of the outcomes.

So the main disagreement I have with Broome comes down to this: on his picture, (rationalizable) preferences that are sensitive to the riskiness of options can only shows up as different specifications of what the outcomes are. They might show up as feelings the agent has about receiving one outcome rather than another – for example, as regret – or they might show up as feelings the agent suffered by getting the outcome in a particular way – for example, as anxiety about getting the outcome as the result of a gamble instead of as a no-fail alternative. Indeed, this strategy makes the same assumption that the "diminishing marginal utility" picture of risk aversion makes: that when an agent appears to care about risk, it must be that he values an outcome in a particular way. There is no room on Broome's picture, as there is on mine, for risk to enter into an agent's feelings about a gamble but *not* about any particular outcome.

Again, I claim that an agent might have reason to be indifferent between two outcomes but to allow them to make different overall contributions to different gambles. Ralph might be indifferent between receiving nothing and receiving nothing when he might have gotten gloves – that is, he might not care about what might have been had he taken a different gamble – because he might not care about particular non-actualized possibilities: the fact that he could have gotten certain prizes, had things turned out differently, does not affect the value of his property in the actual world. It does not affect what he can buy with the money he has, or the amount of pleasure he can get from what he has. In other words, counterfactual money won't pay the bills, or make them harder to pay in the actual world. To put it more concretely, if an agent prefers X to Y, then there is some amount of money he is willing to pay to have X rather than Y (or would be willing to pay if that were possible). But Ralph might not be willing to pay to eliminate possibilities that were never realized. For example, he might not pay any amount of money to trade in "nothing when I might have had gloves" for "nothing when I could not have had gloves" or "$0 when I might have had $1,000,000" for "$0 when I couldn't have had $1,000,000" (if this were possible). We surely cannot blame him for that. And yet, these reasons for not caring about non-actualized possibilities do not undermine his reasons for caring about global properties: the former are all reasons that apply *after* the coin has been flipped. They are considerations about how the values of the outcomes should be affected by other (non-actualized) outcomes. And they apply precisely

[83] And if we try to incorporate them as such, we will get incorrect answers about Ralph's other preferences, e.g. our representation might entail that he prefers gloves without risk to gloves as the result of risk, when, as his other preferences will indicate, this really doesn't make a difference to the value of gloves for him. See my "Risk without Regret" (unpublished paper).

because the other outcomes are *non-actualized*. They cannot apply before the coin has been flipped, when the actualization or non-actualization of the possibilities is not yet known to Ralph. To summarize: caring about risk while all the possibilities are still on the table need not entail experiencing regret in some of these possible occurrences.

Broome's general argument against the existence of counterexamples to STP fails, because there could be reasons for having STP-violating preferences that do not entail differences between outcomes. Again, if it is rational to care about global properties, as I have argued it is, then there are counterexamples to STP that the EU theorist cannot respond to using the individuation strategy and Broome's principle. The failure of Broome's general argument highlighted an important difference between standard treatments of risk-related examples and my treatment, a difference that I have been pointing to all along: I think that aversion to risk should be treated as a feature of the agent's approach to gambles as wholes, not as a feature of how he values particular outcomes. Caring about risk – or, more precisely, evaluating risky gambles in a way that is different from simply averaging the values of the outcomes – need not correspond to caring about any features of the outcomes besides their stated features.

If a decision maker is sensitive to global properties of gambles, then expected utility theory cannot capture his preferences. And if, as I have argued in the previous sections, it is rationally permissible to be globally sensitive in the manner I suggest (i.e., to maximize *risk-weighted* expected utility), then there are rationally permissible sets of preferences that expected utility theory cannot accommodate.

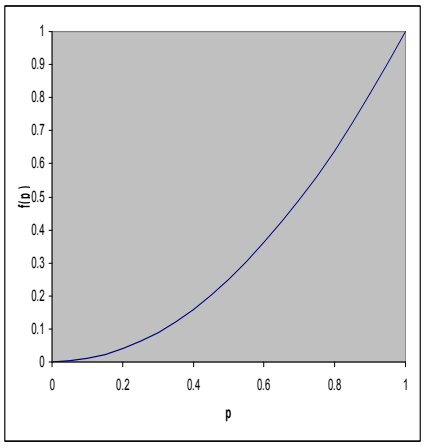
6 Conclusion

Decision theory ingeniously connects belief, desire, and preference; however, the standard theory assumes a particular connection among them that rules out being sensitive to certain global properties of gambles related to their riskiness. There are several sets of preferences – e.g. those of Ralph, Margaret, Watkins, and the Allais agent – that seem reasonable but that expected utility theory is unable to capture. Allais and Hansson, among others, noticed this; and there has been a more recent spate of psychological literature showing how people actually make decisions. However, these deviations from the standard theory are typically thought of as failures of rationality on the part of decision makers (though Allais, notably, didn't think so), and whereas the psychological theories are intended to be purely descriptive, I want to put forth a theory of *rational* preferences among risky gambles. I do not think that all of the noted tendencies of decision makers to deviate from expected utility theory are rational, but I think that preferences arising from attitudes towards risk in particular deserve a more sympathetic

treatment. Indeed, I think that by relaxing a certain assumption and adding another parameter, we can arrive at a theory that is better able to represent the full range of agents with rational attitudes towards risky gambles.

MATHEMATICAL APPENDIX

Appendix A: Example of r-function



Here is an example of a possible r-function. For this agent, the following gambles and sure-thing monetary amounts are equivalent (assuming he values amounts of money under $10 linearly):

Gamble {prize, probability; prize, probability}	Equivalent sure-thing amount
{$0, 0.5; $10, 0.5}	$2.50
{$0, 0.5; $5, 0.5}	$1.25
{$1, 0.5; $9, 0.5}	$3.00
{$0, 0.3; $10, 0.7}	$4.90
{$0, 0.5; $2.50, 0.5}	$0.63
{$2.50, 0.5; $10, 0.5}	$4.38
{$0.63, 0.5; $4.38, 0.5}	$1.57

Appendix B: Ralph's, Margaret's, Watkins's and Allais Preferences are all acceptable on REU

Ralph

	HH	HT	TH	TT
Deal 1	*Elvis stamp*	*Elvis stamp and gloves*	*Nothing*	*Gloves*
Deal 2	*Elvis stamp*	*Elvis stamp*	*Gloves*	*Gloves*

Assuming without loss of generality that the stamp is weakly preferred to the gloves and u(nothing) = 0:
REU(deal 1) = u(nothing) + r(0.75)[u(gloves) – u(nothing)] + r(0.5)[u(stamp) – u(gloves)] +
r(0.25)[u(stamp and gloves) – u(stamp)]
= u(nothing)[1 – r(0.75)] + u(gloves)[r(0.75) – r(0.5)] + u(stamp)[r(0.5) – r(0.25)] + u(both)[r(0.25)]
= u(nothing)[1 – r(0.75)] + u(gloves)[r(0.75) – r(0.5) + r(0.25)] + u(stamp)(r(0.5)), invoking Independence
= u(gloves)[r(0.75) – r(0.5) + r(0.25)] + u(stamp)(r(0.5)), invoking u(nothing) = 0.
= u(gloves)[r(0.75) – r(0.5) + r(0.25)] + u(stamp)(r(0.5))
REU(deal 2) = u(gloves) + r(0.5)[u(stamp) – u(gloves)]
= u(gloves)[1 – r(0.5)] + u(stamp)(r(0.5))
Ralph can prefer deal 2 to deal 1 without contradiction.
E.g., u(stamp) = 1; u(gloves) = 1; $r(p) = p^2$ yields REU(deal 1) = 0.625 and REU(deal 2) = 1

Margaret
u($x) = x with r(p) concave will accommodate Margaret's preferences.
E.g., u($x) = x and $r(p) = p^2$ yields
REU({$0, 0.5; $100, 0.5}) = u($0) + r(0.5)(u($100) – u($0)) = 0 + $(0.5)^2(100 – 0) = 25$.
REU({$50}) = u($50) = 50.
So $50 is preferred to {$0, 0.5; $100, 0.5}.

Watkins
u($x) = x; $r(p) = p^{1.58}$ will accommodate Watkins's preferences:
R1. Whenever faced with a lottery with a 50% chance of one sum of money and a 50% chance of another which is $30 larger, be indifferent between taking that lottery and receiving the lesser prize plus $10 for certain.
REU({x, 0.5; x + 30, 0.5}) = x + r(0.5)(x + 30 – x) ≈ x + 10 = REU({x + 10})
R2. Whenever faced with a lottery with a 50% chance of nothing and a 50% chance of some particular sum, be indifferent between taking this lottery and receiving one-third of that sum for certain
REU({0, 0.5; x, 0.5}) = 0 + r(0.5)(x – 0) ≈ x/3 = REU({x/3})

Allais

Ticket	1	2-11	12 – 100
L_1	*$0*	*$5m*	*$0*
L_2	*$1m*	*$1m*	*$0*
L_3	*$0*	*$5m*	*$1m*
L_4	*$1m*	*$1m*	*$1m*

$L_1 > L_2$ ⇔ u($0) + r(.09)[u($5m) – u($0)] > u($0) + r(.1)[u($1m) – u($0)]
⇔r(.09)[u($5m) – u($0)] > r(.1)[u($1m) – u($0)].
$L_4 > L_3$ ⇔ u($1m) > u($0) + r(.99)[u($1m) – u($0)] + r(.09)[u($5m) – u($1m)].
These two inequalities do not contradict, so an REU maximizer can have the standard Allais preferences.
E.g., u($0) = 0, u($1m) = 5, u($5m) = 7, $r(p) = p^2$ satisfy both of these inequalities.

Works Cited

Allais, Maurice (1953). "Criticisms of the postulates and axioms of the American School." In Rationality in Action: Contemporary Approaches, Paul K. Moser, ed. Cambridge University Press, 1990. (Reprint of 1953 original).

Broome, John (1991). Weighing Goods: Equality, Uncertainty and Time. Blackwell Publishers Ltd.

Broome, John (1999). "Utility." In Ethics out of Economics. Port Chester, NY, USA: Cambridge University Press.

Davidson, Donald (1986). "A Coherence Theory of Truth and Knowledge." In Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson, ed. Ernest Lepore. Blackwell Publishers.

Diamond, Peter (1967). "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: A Comment." *Journal of Political Economy* 75.

Dreier, James (2004). "Decision Theory and Morality." Chapter 9 of Oxford Handbook of Rationality, eds. Alfred R. Mele and Piers Rawling. Oxford University Press.

Ellsberg, Daniel (1962). Risk, Ambiguity and Decision. Routledge, 2001. (Reprint of 1962 original).

Hansson, Bengt (1988). "Risk Aversion as a Problem of Conjoint Measurement." In Decision, Probability, and Utility, eds. Peter Gärdenfors and Nils-Eric Sahlin. Cambridge University Press. Pp 136-158.

Harsanyi, John C. (1977). "On the Rationale of the Bayesian Approach: Comments on Professor Watkins's Paper." In Foundational Problems in the Special Sciences, eds. Butts and Hintikka. D. Reidel Publishing Company, Dordrecht-Holland.

Hurley, Susan (1989). Natural Reasons, Personality, and Polity. Oxford University Press.

Hong, Chew Soo and Peter Wakker (1996). "The Comonotonic Sure-Thing Principle." *Journal of Risk and Uncertainty* 12, pg. 5-27.

Jeffrey, Richard (1965). The Logic of Decision. McGraw Hill.

Joyce, James M. (1999). The Foundations of Causal Decision Theory. Cambridge University Press.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47, pg. 263-291.

Kobberling, Veronika and Peter Wakker (2003). "Preference Foundations for Non-expected Utility: A Generalized and Simplified Technique." *Mathematics of Operations Research* 28, pp 395-423.

Lewis, David (1974). "Radical Interpretation." *Synthese* 23, pp 331-344.

Lewis, David (1983). "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61:4.

MacCrimmon, Kenneth R. (1968). "Descriptive and Normative Implications of Decision Theory." In Risk and Uncertainty, eds. Karl Borch and Jan Mossin. New York: St. Martin's Press.

MacCrimmon, Kenneth R. and Stig Larsson (1979). "Utility Theory: Axioms versus 'Paradoxes." In Expected Utility Hypotheses and the Allais Paradox, eds. Maurice Allais and Ole Hagen. Dordrecht: D. Reidel.

Machina, Mark (1983). "Generalized Expected Utility Analysis and the Nature of Observed Violations of the Independence Axiom." Originally in Foundations of Utility and Risk Theory with Applications, eds. B.P. Stigum and F. Wenstop. Dordrecht: D. Reidel Publishing Company, pp 263-293. Reprinted in Decision, Probability, and Utility, eds. Peter Gärdenfors and Nils-Eric Sahlin. Cambridge University Press, 1988. Pp 215-239.

Machina, Mark (1987). "Problems Solved and Unsolved." *Journal of Economic Perspectives* 1:1, pp 121-154.

Machina, Mark (1991). "Dynamic Consistency and Non-expected Utility." In Foundations of Decision Theory, Michael Bacharach and Susan Hurley, eds. Basil Blackwell.

Maher, Patrick (1993). Betting on Theories. Cambridge: Cambridge University Press.

McClennen, Edward (1983). "Sure-thing doubts." Originally in Foundations of Utility and Risk Theory with Applications, eds. B.P. Stigum and F. Wenstop. Dordrecht: D. Reidel Publishing Company, pp 117-136. Reprinted in Decision, Probability, and Utility, eds. Peter Gärdenfors and Nils-Eric Sahlin. Cambridge University Press, 1988. Pp 166-182.

von Neumann, John and Oskar Morgenstern (1944). Theory of Games and Economic Behavior. Princeton, NJ: Princeton University Press.

Pettit, Philip (2002). "Folk Psychology and Decision Theory," reprinted in Pettit, Rules, Reasons and Norms, Oxford University Press.

Putnam, Hilary (1986). "Rationality in Decision Theory and Ethics." *Critica* 54.

Quiggin, John (1982). "A Theory of Anticipated Utility." *Journal of Economic Behavior and Organization* 3, pg. 323-343.

Ramsey, Frank P. (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. 1999 electronic edition.

Resnik, Michael D. (1987). Choices: An Introduction to Decision Theory. University of Minnesota Press.

Savage, Leonard (1954). The Foundations of Statistics. John Wiley & Sons, Inc.

Savage, Leonard (1972). The Foundations of Statistics. Second edition. New York: Dover Publications, Inc.

Schmeidler, David (1989). "Subjective Probability and Expected Utility without Additivity." *Econometrica* 57, pp 571-587.

Slovic, Paul and Amos Tversky (1974). "Who Accepts Savage's Axiom?" *Behavioral Science* 19, pp 368-373.

Wakker, Peter (1990). "Under Stochastic Dominance Choquet-Expected Utility and Anticipated Utility are Identical." *Theory and Decision* 29:2, pp 199-132.

Wakker, Peter and Amos Tversky (1993). "An Axiomatization of Cumulative Prospect Theory." *Journal of Risk and Uncertainty* 7:7, pp 147-176.

Watkins, J.W.N. (1977). "Towards a Unified Decision Theory: A Non-Bayesian Approach." In Foundational Problems in the Special Sciences, eds. Butts and Hintikka. D. Reidel Publishing Company, Dordrecht-Holland.