

What is Risk Aversion?

April 20, 2016

Abstract

According to the orthodox treatment of risk attitudes in decision theory, such attitudes are explained in terms of the agent's desires about concrete outcomes. The orthodoxy has been criticised both for conflating two types of attitudes and for committing agents to attitudes that do not seem rationally required. To avoid these problems, it has been suggested that we model risk attitudes in terms of a risk function that is independent of the agent's utility and probability functions. The main problem with that approach is that it suggests that attitudes to risk are wholly distinct from people's (non-instrumental) desires. To overcome this problem, we develop a framework where an agent's utility function is partly defined over chance propositions (i.e., propositions describing objective probability distributions), and argue that one should represent different risk attitudes in terms of different forms of the utility function over such propositions.

1 Introduction

In colloquial talk, someone is said to be risk averse if they are disinclined to pursue actions that have a non-negligible chance of resulting in a loss or whose benefits are not guaranteed. This disinclination can be spelled out in a number of different ways. Our focus will be on the one that prevails in much of the literature in economics and decision theory, namely that to be risk averse is to prefer any action A to another with the same expected benefit but with greater variance (canonically called 'a mean-preserving spread' of A). More precisely, let G be any real-valued class of good (e.g. money) and consider lotteries that yield quantities g of G with different probabilities. Then an agent is said to be risk averse with respect to G just in case she prefers g for sure to a (non-trivial) lottery with expectation g . For instance, someone who is risk averse with respect to money will disprefer a gamble yielding either \$0 or \$100 with

equal probability to getting \$50 for sure.

In the orthodox treatment of risk attitudes that prevails in both economics and decision theory, this idea is typically formalised using the expected utility (EU) framework of John von Neumann and Oskar Morgenstern (1944) (hereafter vN-M), though a similar treatment can be given in other frameworks. Within this framework someone who is risk averse with respect to money, say, must, in virtue of the way in which utility is cardinalised, have a concave utility function for money, i.e., must assign diminishing marginal utility to quantities of money. So the orthodox explanations risk aversion with respect to some good G in terms of a particular property of the agent's desires about quantities of G , as captured by the shape of her utility function over G .

This treatment of risk attitudes has been challenged on two different, if related, grounds. First, it has been extensively criticised for failing to distinguish desire attitudes to goods from attitudes to risk itself. Many people feel that because of this failure, the orthodoxy just does not capture the phenomenology of risk attitudes (see e.g. Watkins 1977 and Hansson 1988). For instance, some insist that they evaluate small amounts of money linearly—for example, that the value they attach to \$50 is precisely halfway between those they attach to \$0 and \$100—but nevertheless strictly prefer \$50 to a 50/50 gamble between \$100 and \$0. If this is so, then risk attitudes with respect to a good are not determined by attitudes to quantities of that good the way vN-M suggest. Second, there is a now a large body of empirical evidence suggesting that people exhibit attitudes to risk that cannot be explained within this framework but which are not obviously irrational; most famously in the paradoxes of Allais (1953) and Ellsberg (1961).

In response to these problems with the orthodox treatment of risk, there has been a recent trend towards introducing a special function—a risk function—that, in addition to a probability and a utility function, is used to represent attitudes to risky prospects: most notably in cumulative prospect theory (see Tversky and Wakker 1995 and Wakker 2010) and the rank-dependent utility theory that it draws on (e.g. Quiggin 1982), and in the recent risk-weighted expected utility theory (Buchak 2013). Since these theories account for risk attitudes in terms of the form of the risk function, they can accommodate the intuition that even people who value money linearly might display risk averse behaviour. Moreover, these theories are more permissive than the orthodoxy as to what counts as rational, and thus allow for many of the intuitively rational preferences that the orthodox theory deems irrational. The introduction of

a risk function to account for risk attitudes however raises the question of whether risk attitudes really are a special type of attitude, wholly distinct from non-instrumental desires, as these theories suggest. We shall argue that they are not.

The approach of this paper will differ from this recent trend. We show that it is possible to cardinalise utility without making any assumptions about risk preferences, by extending of Richard Jeffrey’s (1965) decision theory to domains containing *chance propositions*; that is, propositions about objective probability distributions over outcomes of one kind or another. This allows us to model intrinsic attitudes to risks in terms of the form of the agent’s desirability function for chances of goods—thereby respecting the intuition that risk attitudes are a special kind of desire—and to show how such attitudes co-determine, with the agent’s attitudes to the goods themselves, her preferences for risky prospects. In addition to better capturing the phenomenology of risk attitudes than either orthodox expected utility theory or its contemporary rivals, our framework differs from these theories in providing a *unified* explanation of the empirical evidence regarding risk attitudes and so-called ‘ambiguity attitudes’.

2 Risk Attitudes in the vN-M Framework

In the vN-M theory, the utilities of outcomes are determined by the lotteries that the agent is willing to accept. Given lotteries with A and C as their possible outcomes (or ‘prizes’), for instance, we can find the outcomes’ relative utilities by figuring out what chance a lottery L must confer on A for the agent to be indifferent between L and some third outcome B , lying between A and C in the agent’s preference ranking. The basic idea is that your judgment about B , relative to A on the one hand and C on the other, can be measured by the riskiness of the lottery L involving A and C that you deem equally desirable as B . For instance, if you are indifferent between L and B when the chance that L confers on A is $3/4$, then B is three quarters of the way up the utility scale that has C at the bottom and A at the top. This information can be used to determine the *expected* utility of the lottery. If we, say, stipulate that $u(A) = 0$ and $u(C) = 1$, then $u(B) = u(L) = 3/4$. This corresponds to the expected utility of the lottery, since $u(L) = 1/4 \cdot 0 + 3/4 \cdot 1 = 3/4$.

It follows immediately from this way of constructing (cardinal) utilities that the agent must have utilities for lotteries that are linear in the probabilities of the prizes. So if an agent strictly

prefers \$5 to a gamble that will either result in a prize of \$0 or a prize \$10, each prize having a 0.5 chance, then we must, on the vN-M approach, account for this in terms of the agent's attitudes towards the monetary amounts in question, as represented by her utility function. For, as we have seen, agents are assumed to evaluate lotteries by their expected utilities, so the only value that we can adjust in order to account for this person's risk aversion is her utility function over money (assuming that her evaluation of the probability of the prizes is in line with the chances). In particular, we account for this attitude by postulating that money (in the \$0-\$10 range) has decreasing marginal utility to the person (as represented by a concave utility function over money in this range). It is this feature—the assumption that utilities of lotteries are linear in their probabilities—that is at the centre of debate about the adequacy of the orthodox theory.

2.1 Conceptual challenges

A long-standing complaint against the vN-M approach is that it mischaracterises attitudes to risk. Such attitudes, the complaint goes, cannot be explained in terms of attitudes to concrete outcomes. For it seems that two people might evaluate the possible outcomes of a bet in the same way and agree about their probabilities, but nevertheless differ in whether they accept the bet or not, since they have different views about what levels of risk are acceptable (Buchak 2013). Similarly, John Watkins (1977) insists that it is possible for individuals to evaluate monetary outcomes linearly but nevertheless, due to their gambling temperament, turn down bets with a positive monetary expectation. To return to the previous example, it seems a conceptual possibility that a person who has a very strong dislike for gambling would turn down the offer to trade B for the lottery L (which has A and C as possible prizes) except in the special case when L is *certain* to result in A —in which case L is just alternative A —even though she considers B much less desirable than A . In sum, it seems that people's dislike for gambling does not, by itself, tell us much about how they value the gambles' prizes.

The above type of criticism takes as its starting point agents who dislike risk and criticises the vN-M approach for modelling their risk aversion in terms of their attitudes to concrete outcomes. But since the vN-M approach *equates* decreasing marginal utility with risk aversion, it can also be criticised for falsely implying that anyone with a concave utility function over some good is risk averse with respect to that good. Bengt Hansson (1988) for instance tells the

story of a professional gambler who turns down an offer to trade a single copy of a book that he is fond of for an even chance gamble between receiving no copy of the book and three copies. Having been schooled in the vN-M approach, a decision analyst concludes that the gambler must be risk averse. The gambler retorts that this is nonsense; being a professional gambler, he has habituated himself to being risk neutral. The reason he turns down the gamble, he says, is simply that the second and third copy is of almost no worth to him.¹ Since one copy of the book is of great value to him, the gamble he is being offered has an equal chance of resulting in a real loss (losing his single copy) and no real gain (getting two extra copies). And that is the reason he turns down the bet. It is of course possible that someone else might display the same preference between the single book and the gamble due to dislike of risk *per se*. But the gambler turns down the gamble since quantities of the book have decreasing marginal worth to him. These are psychologically distinct reasons that might give rise to the same pattern of preference, and should thus be kept distinct in formal models of practical reasoning.

Decision theorists often respond in one of two ways when confronted with objections like these. The first is to question whether people really can say how they evaluate concrete outcomes without consulting their preferences for risky prospects involving those outcomes; and, correspondingly, whether people really can judge their own risk aversion with respect to some good independently of their preferences between risky prospects involving those goods. But this response does not mitigate the worry that the vN-M approach conflates two distinct attitudes. For we do not need to determine by introspection *precisely* the extent to which we desire outcomes and are willing to accept risk, to be able to justify replies like the gambler's to the decision analyst. Moreover, what Hansson's story illustrates is that, whether or not people can make these judgements by introspection, the vN-M approach equivocates on two phenomena that conceptually and psychologically are very different; on one hand, the decreasing marginal worth of quantities of goods, on the other hand dislike for risk as such. In sum, whether or not people are able to introspect precisely how they evaluate concrete outcomes or what risk they are willing to accept, is orthogonal to the seemingly obvious point that these are two different types of attitudes.

The second type of response that decision theorists and economists typically offer when

¹To make the example particularly plausible, we can assume that the book is of great sentimental value to the gambler but does not have much market value, or, more generally, that the gambler knows that he won't get a price for the second and third copy that matches his evaluation of the first copy.

faced with criticism like that raised above, is to resort to a *formalistic* interpretation of expected utility, according to which the role of expected utility theory is neither to explain nor to guide rational action, but simply to mathematically *represent* rational preferences (or choices). If that is the aim, then as long as we can represent, say, risk averse choice behaviour by postulating a concave utility function, it does not matter that we are equivocating between two conceptually distinct psychological attitudes. In other words, as long as, say, decreasing marginal utility is behaviourally indistinct from what we might call ‘pure’ risk aversion—i.e., aversion to risk *per se*—it does not matter whether or not these are psychologically distinct, since the aim is simply to represent the choice behaviour (see e.g. Harsanyi 1977).

The problem with this response is that we often do want to be able to explain, rather than simply describe, behaviour in terms of the maximisation of a utility function. In other words, we want to be able to say that a person chose an alternative *because* it was the alternative with highest expected utility according to her. Moreover, when using decision theory to make policy recommendations, as is commonly done—or, more generally, when using the theory for decision-making purposes—we need to assume that the utilities on which we base the recommendation exist prior to (and independently of) the choices that the theory recommends. Finally, if we are trying to construct a formal theory of practical reasoning—i.e., trying to characterise how rational agents make choices rather than just mathematically representing choices they have already made—then our theory needs to distinguish the decreasing marginal utility of quantities of goods from aversion to risk as such.

These objections to the orthodox treatment of risk will not get much traction without a demonstration of how utility can be measured without presupposing the vN-M framework. Indeed unless utility can be determined independently of assumptions about the nature of risk attitudes the objections of Watkins and Hansson can be regarded as literally meaningless. But there is a perfectly straightforward response to this worry: *adopt one of the other frameworks for cardinalising utility and test the claims of the vN-M theory within it*. Various such frameworks are already to be found in the decision-theoretic literature, including the aforementioned cumulative prospect theory and risk-weighted expected utility. But, for reasons that we will explain later on, we consider these frameworks to have their own problems, both conceptual and empirical. So instead we will make use of Bayesian² decision theory to cardinalise utility;

²We call a decision theory ‘Bayesian’ if the probabilities that go into the expectation whose value rational agents maximise are the agent’s own subjective probabilities. Thus Savage’s (1954) and Jeffrey’s (1965) decision

	<i>Gains</i>	<i>Losses</i>
<i>Low probability</i>	Risk Seeking	Risk Aversion
<i>High probability</i>	Risk Aversion	Risk Seeking

Table 1: Fourfold Pattern of Risk Preferences

in particular an extended version of the variant developed by Richard Jeffrey (1965) that has the virtue of not implying that the value of a lottery is linear in its probabilities.

2.2 Empirical challenges

The orthodox treatment of risk attitudes faces two distinct types of challenges from the growing body of evidence regarding people’s actual choices in situations of risk and uncertainty.³ Firstly, there is much evidence that people exhibit risk attitudes in their choices that cannot be reconciled with expected utility theory and in particular that their preferences between risky prospects are not linear in the chances of the outcomes. In Kahneman and Tversky’s (1979) famous study, for instance, they report the fourfold pattern of attitudes for simple lotteries of the form ‘x chance of \$y’ displayed in Table 1, obtained by determining the agents’ dollar prices for these lotteries. The pattern of risk *averse* behaviour when it comes to lotteries with high probability of monetary gains or low probability of losses, together with risk *seeking* behaviour for lotteries with low probability of monetary gain or high probability of losses, cannot be reconciled with vN-M’s expected utility theory no matter what utility function is attributed to subjects. This has lead most decision theorists to conclude that expected utility theory is not descriptively adequate as a theory of choice amongst risky prospects.

It is possible that the deviation from the predictions of the orthodox theory exhibited by this pattern of choice is to be attributed to irrationality on the part of the deviating subjects. And indeed the focus of interest in the decision-theoretic literature has been on the implications of these results for descriptive decision theory, with little in the way of a consensus emerging on their normative implications. But some instances of these risk preferences do not seem irrational. The most famous example of this is the so-called *Allais Paradox*, originally introduced by Maurice Allais (1953). The ‘paradox’ is generated by comparing people’s preferences over two pairs of lotteries similar to those given in Table 2. The lotteries consist in tickets being

theories are Bayesian, but von Neumann and Morgentern’s (1944) is not.

³We will follow the convention of calling decision situations where the relevant outcomes have known and objective probabilities situations of *risk*, and we will call decision situations where the decision-maker lacks such knowledge situations of *uncertainty*.

randomly drawn, determining the prize of each lottery (for instance, lottery L_1 results in a prize of 5 million dollars if one of the tickets numbered 2-11 is drawn).

	1	2 – 11	12 – 100
L_1	\$0	\$5m	\$1m
L_2	\$1m	\$1m	\$1m
L_3	\$0	\$5m	\$0
L_4	\$1m	\$1m	\$0

Table 2: Allais' paradox

In this situation, many people strictly prefer L_2 over L_1 but also L_3 over L_4 , a pair of preferences we will call *the Allais preference*. According to the orthodox theory, this pattern of preference is irrational, since there is no way to assign utilities to the prizes on offer such that L_2 gets a higher expected utility than L_1 and L_3 gets a higher expected utility than L_4 . In other words, the Allais preference cannot be represented as maximising expected utility (which according to the orthodox picture implies that it is irrational).

Here is an explanation of why the reasoning underlying the Allais-preference is inconsistent with EU theory. People with this preference find that the value of decreasing the risk of \$0 (from a base-line of \$1m) from 0.01 to 0 exceeds the value of a 0.1 chance at \$5m rather than \$1m. And that is why they prefer L_2 to L_1 . However, the same marginal decrease in the risk of \$0 does not exceed the value of a 0.1 chance at \$5m rather than \$1m when the decrease is from 0.9 to 0.89. And that is why they prefer L_4 to L_3 . (In other words, people typically consider a reduction in the risk of winning nothing from very unlikely to impossible to be more important than the same absolute decrease in risk from quite likely to only slightly less likely.) But in both cases, people are comparing a 0.01 chance of getting \$0 rather than \$1m with a 0.1 chance of getting \$5m rather than \$1m. So what they are comparing on both occasions, according to vN-M's theory, is $0.01[u(\$1m) - u(\$0)]$ with $0.1[u(\$5m) - u(\$1m)]$ (where $u(x)$ denotes the utility of x). So if this reduction in risk of ending up with \$0 is worth foregoing a 0.1 chance at \$5m when comparing L_1 and L_2 , it should also, on the orthodox story, be worth it when comparing L_3 and L_4 .

The second type of empirical challenge that the orthodox EU theory is faced with relates to choices that agents make in contexts characterised by *both* (subjective) uncertainty and (objective) risk. In particular, the orthodoxy is unable to account for the phenomenon of

ambiguity aversion, a pattern of preference that is typically exhibited in the famous *Ellsberg Paradox*, but which can be much more simply explained with the following example.⁴

Suppose you have in front of you a coin, C_1 , that you know to be perfectly symmetric, and that you know to have been tossed a great number of times and has come up heads exactly as many times as it has come up tails. More generally, suppose that you possess the best possible evidence for the coin being unbiased. Here are two questions: (1) How confident⁵ are you that C_1 will come up heads on its next toss? (2) How much would you be willing to pay for a bet that pays you \$10 if C_1 lands heads on its next toss but pays nothing otherwise?

Now suppose instead that you have in front of you a coin, C_2 , that you know to be either double headed or double tailed, but you don't know which. Here are again two analogous questions: (1') How confident are you that C_2 will come up heads on its next toss? (2') How much would you be willing to pay for a bet that pays you \$10 if C_2 lands heads on its next toss but pays nothing otherwise?

Most people seem to use something like the *Principle of Insufficient Reason* when answering questions like (1') (see e.g. Voorhoeve, et al., 2012). Since they have no more reason for thinking that the coin will come up heads than tails, they are equally confident in these two possibilities. But since these possibilities exhaust the possibility space, they should believe to degree 0.5 that the second coin comes up heads. But that is, of course, the same degree of belief as they should have in the proposition that the first coin comes up heads (assuming something like Lewis' 1980 *Principal Principle*). There is, of course, an important difference between their judgements about the two coins, as we will discuss in more detail later on. In the first case, they are pretty certain that the coin has an (objective) chance of one-half of coming up heads on the next toss; in the latter case they are not. But in both cases, they are equally confident that the coin will come up heads on the next toss as that it will come up tails.

What about questions (2) and (2')? A number of experimental results on Ellsberg-type decision problems show that people tend to be what is called ambiguity averse, meaning that they prefer prospects with known chances of outcomes to one's with unknown chances (see e.g. Wakker 2010). In the example under discussion, ambiguity aversion translates into a preference

⁴This version of the paradox assumes that we can enquire about people's degrees of belief without consulting their preferences between risky prospects. Those who adhere to the formalistic interpretation of Bayesian expected utility theory—and behaviourists more generally—are unlikely to accept that assumption. So it is worth noting that the original choice problem described by Ellsberg (1961) does not require that assumption.

⁵We will use 'confidence' and 'degree of belief' interchangeably.

for a bet on C_1 over a bet on C_2 and hence a willingness to pay more for the first bet than the second.

Now the above answers may seem to create a problem for Bayesian decision theory. Since the possible prizes are the same for the two bets, standard applications of the theory imply that people should be willing to pay more for the first bet than the second only if they are more confident that they will get the prize (the \$10) if they accept the first bet than if they accept the second bet. But they are not: they are confident to degree one-half of getting the \$10 in both cases. In response, orthodox Bayesians might try to argue that people who are willing to pay more for the first bet than the second have made some mistake in their instrumental reasoning. But that seems implausible. Ambiguity aversion, even in these very simple set-ups, is a robust phenomenon, making it unlikely that people are simply making a mistake.

In fact however the above analysis ignores the important difference between the two cases. In the first case, a bet on heads amount to accepting a lottery which confers a chance of one-half on the prize. In the second case a bet on heads is an action which yields the prize with chance one or with chance zero depending on whether the coin is two-headed or two-tailed. But this difference between the two cases is irrelevant if the vN-M theory is correct. The upshot is that ambiguity aversion, being a phenomenon that arises when both subjective and objective uncertainty is present, raises a challenge to the combination of the Bayesian theory of rational preference under uncertainty and the vN-M theory. Most of the literature on ambiguity aversion is based on the assumption that the vN-M theory is correct and hence draws the conclusion that ambiguity aversion is inconsistent with Bayesian rationality. We will take the contrary view, arguing that ambiguity aversion is a permissible attitude to spreads of chances that is perfectly consistent with the kind of Bayesian framework in which we will work, but inconsistent with the vN-M theory.

Although the challenges presented by the empirical evidence concerning attitudes to risk seem quite different from those presented by attitudes to ambiguity, we will see that the reason why these two types of attitudes generate trouble for orthodox expected utility theory is much same; namely, the theory's narrow conception of the (dis)value of risks and chances. As we show in section 4, it is possible to account for both types of attitudes in a decision theory whose value function is defined in part over a set of chance propositions. Moreover, this makes it possible to represent an Allais-type preference and an ambiguity averse preference as maximising the

value of the same desirability function, which means that we can offer a unified explanation of the two (types of) empirical observations that have posed the greatest challenge to orthodox decision theory. But first, let us discuss a recent and quite influential alternative to expected utility theory, and explain why we think that this alternative does not do justice to ordinary risk attitudes either.

3 Risk-Weighted Expected Utility Theory

Recently several authors have constructed non-expected utility theories that introduce a risk function to represent people's risk attitudes, with the aim, first, to formally capture the intuition that risk attitudes with respect to some good need not be determined by how people evaluate quantities of that good; and, second, to make it possible to represent preferences like Allais' as maximising agents' value functions. We will focus on Lara Buchak's (2013) recent *risk-weighted* expected utility (REU) theory as a representative of these non-expected utility theories, but our argument equally applies to (normative versions of) the theories on which Buchak's theory is based (such as the rank-dependent utility theory). The simplest way to explain the risk-weighted expected utility theory is by comparing it to classical (Bayesian) expected utility theory. We do so in the next subsection, and raise two objections to REU theory in the subsection after that.

3.1 REU versus EU

One of the main differences between REU theory and more traditional Bayesian decision theory concerns how many variables are determined by the agent we are trying to model, and, correspondingly, how many functions we use to represent her mental attitudes. The orthodox theory leaves it up to the agent to decide two things: first, the values of the possible consequences of the acts at the agent's disposal, as represented by her utility function; second, the probabilities of the different contingencies that determine which of these consequences are realised when each act is performed, as represented by her subjective probability function. In addition, REU theory leaves it up to the agents to decide how to aggregate the values of different possible outcomes of an alternative in order to evaluate the overall value of the alternative, and represents this by a *risk function*. The idea is that the form of an agent's aggregation will depend on

how she trades-off chances for good outcomes against risks of bad outcomes. So whereas EU theory models rational agents as maximising expected utility relative to a pair of utility and probability functions, REU theory models rational agents as maximising *risk-weighted* expected utility relative to a triple of utility, probability and risk functions.

To make the discussion that follows more precise, let r be a (non-decreasing) risk function, satisfying the constraint that $0 \leq r(p) \leq 1$ and $r(0) = 0$, $r(1) = 1$. The function is intuitively to be understood as a weighing function on probabilities, whose purpose is to discount or inflate, in accordance with the agent’s attitudes to risk, the probability of attaining more than the minimum that an alternative guarantees. Let s_i denote a state of the world, and $f(s_i)$ the outcome of act f when state s_i is actual. u is a utility function on outcomes and P a probability function on states. Now the value of f , according to Buchak, is given by its risk weighted expected utility, which, she argues, rational preferences maximise:

$$REU(f) = \sum_i u(f(s_i)) \cdot r(P(s_i)) \tag{1}$$

Risk-weighted expected utility theory can allow for the possibility that two individuals with the same beliefs and the same desires over risk-free outcomes differ in their evaluation of risky prospects. Suppose both Ann’s and Bob’s beliefs can be represented by the same probability function and that they evaluate monetary outcomes in the same way. Nevertheless, Ann is willing to pay up to and including \$5 for an even chance gamble that either results in her winning \$10 or nothing, whereas Bob is only willing to pay \$3 for the same gamble. As we have seen, two people cannot differ in this way, according to orthodox EU theory: given the difference between Ann’s and Bob’s attitudes to these gambles, they must either have different beliefs or disagree about the relative values of the prizes on offer. In contrast, risk-weighted expected utility theory can account for the above difference between Ann and Bob’s attitudes to gambles, without postulating different probability or utility functions, by assuming that Ann’s risk function is linear ($r(p) = p$) while Bob’s is convex (in particular, $r(0.5) < 0.5$).

Risk-weighted expected utility theory can potentially account for the Allais preference in a similar way. Given a linearity assumption which Buchak (2013: fn. 39) implicitly makes—but which we will not—the Allais preference can be represented as maximising risk-weighted expected utility whenever the risk function is convex, in particular, when the difference between

$r(1)$ and $r(0.99)$ is greater than the difference between $r(0.11)$ and $r(0.10)$. The intuitive explanation for this, you may recall, is that a 0.01 probability difference counts more heavily in the agent’s decision-making when it means that a prize becomes certain as opposed to almost certain, than when it means that a prize becomes only slightly more probable but still quite unlikely.

3.2 Problems with REU theory

In this section we raise two problems for REU theory. The first is that it cannot (nor is it meant to) account for the Ellsberg preferences. Our simplified version of the Ellsberg paradox suffices to illustrate this. Recall that we are assuming that a person is equally confident that C_1 and C_2 will come up heads on their next tosses; she is offered a bet on each coin that pays her \$10 if it comes up heads on the next toss but nothing otherwise; and she is willing to pay a higher price for the bet on the first coin than the second, since only in the former case is she confident that the coin is unbiased. The risk-weighted expected utility of *both* bets is equal to $r(0.5) \cdot u(\$10) + r(0.5) \cdot u(\$0)$. Hence, REU theory cannot make sense of the willingness to pay more for one of these bets than the other.⁶

The second problem with risk-weighted expected utility theory is that even in those cases where it is consistent with risk averse preferences such as Allais’, it mischaracterises the psychology of risk attitudes. This is an especially grave problem for a theory like Buchak’s, whose benefit compared to orthodox expected utility theory is partly meant to be that it better fits the phenomenology of risk attitudes. The problem essentially consists in the fact that risk attitudes are, according to REU theory, primitive mental attitudes, distinct from both desires and beliefs. The risk function, r , is logically independent of both the probability function, P , and, more importantly for the present argument, also independent of the utility function, u . This is not an accident. As Buchak (2013: 53-54) explains:

The utility function is supposed to represent desire ... and the probability function is supposed to represent belief ... We try to make beliefs ‘fit the world,’ and we try

⁶Another problem with EU theory that REU theory cannot solve, unlike the theory developed in the next section, is the Diamond (1967) ‘paradox’, which is based on EU theory’s inability to account for the intuition that sometimes it is valuable to give people a chance at a good even if they do not end up receiving the good. See Stefánsson and Bradley (2015) for an explanation of how the framework discussed in the next section is partly motivated by the problem raised by Diamond, and Stefánsson (2015: sec. 4) for a demonstration of REU theory’s inability to solve it.

to make the world fit our desires. But the risk function is neither of these things: it does not quantify how we see the world—it does not, for example, measure the strength of an agent’s belief that things will go well or poorly for him—and it does not describe how we would like the world to be. It is not a belief about how much risk one should tolerate, nor is it a desire for more or less risk. The risk function corresponds neither to beliefs nor desires. Instead, it measures how an agent structures the potential realizations of some of his aims.

We are happy to grant Buchak the claim that the attitudes that the risk function is meant to represent are not beliefs. But we find it hard to understand her view that these risk attitudes are not a special kind of desire, especially if we accept her (quite standard) characterisation of desire as the type of attitude to which we try to make the world fit.⁷ Recall Bob, who values money linearly but is nevertheless risk averse; for instance, prefers \$5 for sure to a gamble whose monetary expectation is \$5. Surely, any risk function that reflects this fact about Bob partly describes how he would like the world to be. In particular, he would rather like the world to be such that he holds \$5 than a bet with the same monetary expectation, and his risk function reflects this wish. Moreover, Bob will, if instrumentally rational, try what he can to make the world fit this attitude, for instance by not accepting certain bets, and by hedging the risks he exposed to. So risk attitudes are attitudes to which we try to fit the world.

More generally, people who are risk averse have different views than the risk neutral (or risk loving) about how outcomes should be distributed across the possibility space, as Buchak herself points out (2013: 29). Informally put, risk averse people prefer goods (*including chances*) to be spread evenly over the possibility space, such that they are guaranteed to get something that is not too bad no matter what the world happens to be like. Risk loving individuals, on the other hand, prefer goods to be more concentrated, such that if a state favourable to them turns out to be actual, they get lots of the good in question. Again, these different attitudes will manifest themselves in terms of different ways of trying to arrange the world; the risk averse trying to arrange the world such that goods are evenly spread across the space of possibilities (e.g. by hedging their bets), but the risk neutral trying to have them more concentrated (e.g. by accepting risky bets).

Buchak emphasises in various places that she takes the risk function to represent how an

⁷Some of these arguments against Buchak’s view can be found in Stefánsson (2014).

individual “structures the potential realizations of some of his aims” (ibid.), which she takes to be incompatible with seeing the risk function as representing part of an agent’s desires. But it is unclear why these are incompatible interpretations. For it seems that the risk function represents how an agent *would want to* (and will if she can) structure the realization of her aims. Other things being equal, risk averse people will want to, and will try to, realise their aims as safely as they can, even if that reduces the expected realisation of their aims. For instance, they will, other things being equal, want to (and try to) structure the gambles they hold in such a way to spread goods equally rather than unequally across the possibility space. Doesn’t this mean that they *desire* their gambles and other prospects to be structured in this way rather than in a more risky way?

In sum, it seems to us that attitudes to risk are simply a special kind of desire, rather than a primitive mental attitude on a par with beliefs and desires. But they are not desires about concrete outcomes, as already discussed. So in that respect we agree with Buchak’s criticism of orthodox expected utility theory. Instead, they are desires about *chance distributions*. That is, risk averse people want chances to be distributed one way, risk neutral and risk loving people in other ways. In the next section we will make this suggestion more precise, by presenting a decision theoretic framework where people’s value functions are partly defined over propositions about chances. We will show how this framework, first, respects the intuition that risk attitudes are not about concrete outcomes but are still a special kind of desire, and, second, makes it possible to represent both Ellsberg- and Allais-type preferences as maximising the expectation of agents’ value functions.

4 Risk Attitudes in the Jeffrey Framework

Our aim in this section is to present a framework in which preferences for risky prospects can be cardinalised with only minimal assumptions about the properties of such preferences and in particular without assuming that they are linear in chances. To do so we build on Richard Jeffrey’s version of Bayesian decision theory. His theory has two great advantages in this context, compared to the rival Bayesian theory of Leonard Savage (1954). First, since the objects of desire in Jeffrey’s theory are propositions, it is considerably more natural to extend the theory to allow for conative attitudes to chance distributions than to similarly extend

Savage’s theory, where the objects of desire are interpreted as acts and formally modelled as functions from the state space to the set of consequences. Second, Jeffrey’s theory depends on much weaker assumptions than Savage’s, which enables us to evaluate proposals for constraints on rational attitudes to risks and chances without taking too much for granted.

4.1 Linearity, Chance Neutrality and risk aversion

In Jeffrey’s theory the degrees of belief of a rational agent are measured by a subjective probability function, P , on a Boolean algebra of propositions, Ω . Her degrees of desire, however, are measured by a corresponding desirability function, V , defined on the same algebra but with the logically contradictory proposition \perp removed and satisfying for all $A, B \in \Omega - \{\perp\}$:

Desirability: If $A \wedge B = \perp$, and $P(A \vee B) \neq 0$, then:

$$V(A \vee B) = \frac{V(A).P(A) + V(B).P(B)}{P(A \vee B)}$$

Necessary and sufficient conditions for preferences to be represented by such a pair of functions, P and V , were established by Ethan Bolker (1966). None of these conditions have anything special to say about preferences for risky prospects. Indeed there are no lotteries in the basic Jeffrey-Bolker framework, so it is not possible to model risk preferences within it (let alone cardinalise utility on the basis of them). This is a limitation that we now need to address. We do so by explicitly introducing propositions about chances and then identifying lotteries with conjunctions of such propositions.

Let \mathcal{Z} be a Boolean subalgebra of the background Boolean algebra Ω ; intuitively \mathcal{Z} contains those propositions to which it is meaningful to ascribe chances.⁸ Let $\Pi = \{ch\}$ be the set of all probability functions on \mathcal{Z} and let $\Delta = \wp(\Pi)$ be the set of all subsets of Π . The elements of Δ serve here as what we will call *chance propositions*. In particular, for any $X \in \mathcal{Z}$, and $x \in [0, 1]$, let $Ch(X) = x$ denote the chance proposition defined by $\{ch \in \Delta : ch(X) = x\}$. Intuitively $Ch(X) = x$ is the proposition that the chance of X is x (and the chance of $\neg X$ is $1 - x$). Let $\mathcal{X} = \{X_1, \dots, X_n\}$ be an n -fold partition of \mathcal{Z} , with the X_i being the prospects that constitute the various possible ‘prizes’ of a lottery or, more generally, outcomes

⁸Canonically we take the base propositions to be sets of possible worlds, but nothing hangs on this particular treatment of them.

of some stochastic process. Let $\bigcap_{i=1}^n (Ch(X_i) = x_i)$ denote the conjunction of the n propositions $Ch(X_1) = x_1, Ch(X_2) = x_2, \dots,$ and $Ch(X_n) = x_n$, where the x_i are such that $\sum_{i=1}^n x_i = 1$. A proposition $\bigcap_{i=1}^n (Ch(X_i) = x_i)$ thus expresses the chances of realising each of the X_i , thereby serving as the propositional equivalent, in this framework, of the lotteries that are the basic objects of choice in the orthodox (vN-M) theory of decision-making under risk.

The focus of our interest is the product set $\Gamma = \mathcal{Z} \times \Delta$ whose elements are combinations of factual and chance prospects. Since Γ forms a Boolean algebra we can simply apply Bolker's theorem to establish, for preferences over the propositions in $\Gamma - \{\perp\}$ satisfying the Bolker axioms, the existence of a probability function P and desirability function V , respectively on Γ and $\Gamma - \{\perp\}$, measuring the agent's degrees of belief in, and desire for, the propositions contained in these sets; including, of course, propositions concerning chances. We thus have a framework in which it is meaningful to ask what the relationship is between agents' attitudes to concrete goods and their attitudes to chances of such goods, including lotteries over them. And in particular whether or not agent's preferences for lotteries must generally satisfy the requirements of the vN-M theory or not.

We can give an immediate answer to the latter question. As we have seen, the vN-M theory postulates that agents' utilities for lotteries are linear in the chances. This is captured in our framework by the following condition on preferences for chance propositions:

Linearity: For any n -fold partition \mathcal{X} of Γ and set $\{x_i\}$ such that $x_i \in [0, 1]$ and $\sum_{i=1}^n x_i = 1$:

$$V\left(\bigcap_{i=1}^n (Ch(X_i) = x_i)\right) = \sum_i x_i \cdot V(X_i)$$

Linearity says that the desirability of any lottery is a sum of the desirabilities of the lottery's prizes weighted by the chances accorded to them by the lottery. Informally, we can think of it as capturing the idea that chances do not matter intrinsically to the agent; they matter only instrumentally, as means to the attainment of the prizes that they are chances of. More exactly, as shown in Stefánsson and Bradley (2015: Theorem 1), Linearity encodes the neutrality of chances, the idea that one should not care about the chance of a (maximally specific) outcome once one knows whether or not the outcome has been realised. Formally (and adopting the convention that the status quo has desirability zero):

Chance Neutrality: $V(Ch(X) = x|X) = 0$

So the question of whether or not an agent's risk preferences within the extended framework must conform with the vN-M theory boils down to that of whether it is rationally permissible or not to attach any (dis)value to chances over and above the extent to which they make various goods of which they are the chances of more or less likely. It seems to us that Chance Neutrality makes a substantial value claim and is not a mere consistency condition on desire. Hence, on the decision theoretic conception of rationality as consistency, people are rationally permitted to like (and dislike) chances non-instrumentally.

We will not pursue this normative issue any further however as it has been argued at length elsewhere that Linearity and Chance Neutrality are not requirements of rationality (see Stefánsson and Bradley 2015). Instead let us return to the opening observation that dislike of risk *per se*, rational or otherwise, is psychologically very different from the decreasing marginal desirability of quantities of concrete goods, even though the two phenomena may give rise to the same choice behaviour. Can we make sense of this observation within our framework? To do so we must show how we can distinguish the two kinds of attitudes and then show how these attitudes relate to risk aversion as characterised behaviourally by a preference for a lottery over any mean-preserving spread of it. Let's consider each task in turn.

4.2 Distinguishing risk attitudes

Our central thesis is that an agent's like or dislike of risk involving a good is captured by the properties of her desirability function for chances of this good. Consider a lottery that pays \$100 with probability one-half and nothing otherwise and an agent whose desirabilities for modest amounts of money are linear in quantities of it. If this agent attaches a desirability to the half-chance of winning the \$100 equal to the desirability she attaches to winning \$50, then we can say that she is neutral with regard to the risk of (not) winning the \$100. But if the desirability of the half-chance of winning the \$100 is less than that of \$50, then she must value the half chance of \$100 at less than half the value of the \$100. So in this case the agent will displays behavioural risk aversion (she will prefer the \$50 to the lottery), not because of her attitude to quantities of money (which were assumed to be linear), but because of her attitude to the risk *per se* of (not) winning the \$100. So unlike the vN-M theory, our framework can distinguish between these two quite different motivations for the same choice behaviour.

In general, in our framework, an agent's preferences amongst lotteries will depend not just on

her desirability function for quantities of the good at stake, but also on her desirability function for chances of these goods (indeed on the relationship between the two). To substantiate this claim let us restrict attention to a class of simple lotteries identified by propositions of the form ‘The chance of \$100 is x ’ where, of course, $x \in [0, 1]$. In Figure 1, we plot some example desirability functions for these propositions against values for x (the chances). For convenience, we set the desirability of a zero chance of \$100 to zero and the desirability of \$100 for sure to 1. If the vN-M theory is correct, then the relationship between chances of \$100 and their desirability is linear and so the graph will be a straight line. On the other hand, if the chances have diminishing marginal desirability then the graph will be concave. But there are other possibilities: chances could have increasing marginal desirability, or a combination of increasing and decreasing marginal desirability at different chances, as illustrated by the snake-shaped curve plotted in Figure 1. Evidently, it is an empirical matter as to what attitudes to chances agents actually display and there is no *a priori* reason why the graph should have one shape rather than another (and no shape is imposed by the adoption of our framework).

Figure 1: Candidate Desirability Functions for Chances of \$100

To see how these curves capture the agent’s attitudes to risk, let us define a risk function R on chances for an agent from her degrees of desire by setting $R(x)$ equal to $V(\$y)$ where $\$y$ is the amount of money such that the agent is indifferent between getting it with certainty and getting \$100 with chance x . So:

$$V(x, \$100) = V(\$y) = R(x)$$

Now the function R will behave much like the risk functions deployed in cumulative prospect theory and risk-weighted EU theory since, given our choice of scaling of the desirability function, $V(x, \$100) = R(x) \cdot V(\$100)$; i.e., the desirability of some chance of \$100 will equal the risk-weighted desirability of \$100. (Nothing depends on this choice of scaling, but without it a somewhat more complicated definition of R would be required.) So we can interpret the graphs in Figure 1 as candidate risk curves, representing the agent’s attitude to risk as such, with the linear one being the vN-M risk curve and the snake-shaped one being the curve postulated by cumulative prospect theory. Crucially, however, the risk curves, so defined, are features of the

agent’s desires and not some distinct attitude. More precisely, these risk curves are determined by the relationship between the agent’s desires for concrete goods and her desires for the chances of these goods. Hence, there is no reason to expect that properties of an agent’s attitudes to the chances of one good, say money, will be the same as her attitudes to the chances of another, say health. A professional gambler who bets only to maximise expected monetary value when in the casino, may be extremely averse to taking risks with his health.

It is the way in which the attitudes that agents take to goods combine with the attitudes they take to chances of goods that gives rise to behavioural risk aversion as initially defined, namely as a preference for lotteries over mean preserving spreads of them. For instance, diminishing marginal desirability of quantities of a good combined with constant marginal desirability of chances of the good—the only attitude to chances that the orthodox theory allows—gives rise to such risk averse behaviour. But so does constant marginal desirability of quantities of the good combined with *increasing* marginal desirability of the chances. This case is illustrated in Figure 2. Similarly, an agent may display risk neutrality in their choices (i.e., indifference between a lottery and any mean-preserving spread of it) because they assign increasing marginal desirability to both quantities of the good and chances of it, or because they assign constant marginal desirability to both, or because they assign decreasing marginal desirability to both!

Figure 2: Linear Desirability for Quantities of Money and Increasing Marginal Desirability for Chances of \$100

4.3 Ambiguity and the fourfold pattern

Let us now turn to explanation of the two empirical phenomena that standard expected utility theory has such difficulty accommodating. The explanation we offer of an agent’s ambiguity attitudes towards some good is very straightforward; they are simply the reflection of the shape of her desirability function for the chances of a good. In particular, ambiguity aversion with respect to actions with consequences that are chances of monetary prizes reflects the diminishing marginal desirability of the chances of money. So on our account, ambiguity attitudes to goods are simply attitudes to the chances of these goods.

Consider again the example of bets on the two different coins that we gave earlier. A bet that the coin that is known to be fair will land heads has a ‘sure’ consequence of a chance

of one-half of winning \$10. Hence, the desirability of this bet is equal to the desirability of a half chance of winning \$10. But the corresponding bet on the other coin has different consequences in different states of the world: if the coin is two-tailed then the bet has no chance of delivering the \$10, but if it is two-headed then it is certain to result in a win of \$10. Its desirability, assuming that the agent assigns equal probability to both possibilities (as revealed in an indifference between betting on heads and betting on tails), is the average of the desirability of no chance of winning \$10 and the desirability of winning \$10 for sure. Now if these chances have diminishing marginal desirabilities, then, by definition, the difference between a half chance and a zero chance of winning is greater than that between certainty of winning and a half chance of doing so. So then rationality requires that the agent prefer the bet on the fair coin, i.e., that she displays ambiguity aversion.

The same explanation can be offered of the pattern of preferences exhibited in the Ellsberg paradox. Far from being inconsistent with Bayesian rationality, these preferences are, as Bradley (2016) shows, required of Bayesian agents who attach diminishing marginal desirability to chances of monetary outcomes. More generally, an agent with diminishing marginal desirabilities for the chances of some good will exhibit the kind of preference for hedging chances that is characteristic of ambiguity aversion.

The explanation of the fourfold pattern of risk preferences (in particular, the Allais preference) reported by Kahneman and Tversky (1979) is equally straight-forward. These results should be read as reporting the exchange rates between quantities of money and chances of winning a fixed amount of money induced by the agents' degrees of desire for these two types of good. So we can infer from them what sorts of relationships must hold between these desires. For instance, they report for lotteries yielding \$100 with different chances, the following median cash equivalents: \$14 for chance 0.05, \$25 for chance 0.25, \$36 for chance 0.5, \$52 for chance 0.75, and \$78 for chance 0.95. These exchange ratios reflect the relationship between the agents' attitudes to monetary amounts in the range \$0 to \$100 and their attitudes to different chances of \$100. What cannot be determined from such data is what the shape of each desirability function is. We can however conclude that the subjects value gains in chances of \$100 more highly than the corresponding expected gains in monetary amounts when both the absolute chances and monetary amounts are small, but the other way around when both are large. So agents will pay much larger sums of money for a gain in chances when these chances are small

than when they are large.

One way in which these constraints could be satisfied is if agents have desirability functions for money that are roughly linear in quantities and snake-shaped desirability functions for chances, concave for small probabilities and convex for the very high ones. This is what is predicted by cumulative prospect theory. Such a postulate implies however that a Bayesian agent will exhibit ambiguity seeking preferences in situations involving lotteries with high chances of winning \$100. The empirical evidence offers little support for this implication; indeed it is not consistent with the ambiguity averse patterns of preference frequently observed in the Ellsberg paradox.

Figure 3: Desirability Function for Quantities of Money and Chances of \$100

In contrast, the fourfold pattern of risk preferences is perfectly consistent with ambiguity aversion in our framework. An agent will display both the fourfold pattern and ambiguity aversion when the desirability function is concave over both quantities of money and chances, but relatively less so over low chances of some amount than small percentages of the amount and relatively more so over high chances of the amount than over large percentages of the amount. To illustrate, suppose that agents have the desirability functions over monetary gains and chances of \$100 depicted in Figure 3. Note that although they have diminishing marginal desirabilities for both, the shapes of the two functions are different, with the desirabilities of chances initially rising more rapidly than those of the monetary amounts, but less rapidly later on. Such agents would display precisely the risk seeking preferences at low chances and risk aversion at high chances that Tversky and Kahneman report. For instance, they would be willing to pay more than \$1 to achieve a 0.01 chance at a \$100 dollar prize when their chances of the prize are 0, and would also be willing to pay more than \$1 to avoid a 0.01 drop in the chance of \$100 when their chances are 1 (i.e., when they have already secured the prize). They would also of course exhibit ambiguity aversion because of the concave shape of the desirability function for chances.

5 Conclusion

The orthodox treatment of risk attitudes in decision theory seems both conceptually and empirically inadequate. Conceptually because it fails to distinguish attitudes to concrete goods from attitudes to risks regarding these goods; empirically because it neither offers a satisfactory explanation of the fourfold pattern of risk behaviour observed in choice experiments, most famously in the Allais paradox, nor of ambiguity attitudes observed in setups such as the Ellsberg paradox. We have offered a framework in which it is both possible and natural to distinguish attitudes to concrete goods from risk attitudes, and shown how these two types of attitudes can combine to determine an agent's choices in a way that is consistent with both the fourfold pattern and with ambiguity aversion. In particular, an agent with diminishing marginal desirabilities for chances of money will display ambiguity aversion in her choices in conditions involving both subjective and objective uncertainty. Furthermore, if her desirabilities for chances of money increase more rapidly than those of the corresponding monetary quantities at both low chances and high chances, but less rapidly in between, then she will display the fourfold pattern of risk preferences typically observed.

References

- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine. *Econometrica* 21(4), 503–546.
- Bolker, E. D. (1966). Functions resembling quotients of measures. *Transactions of the American Mathematical Society* 124(2), 292–312.
- Bradley, R. (2016). Ellsberg's paradox and the value of chances. *Economics and Philosophy* (forthcoming).
- Buchak, L. (2013). *Risk and Rationality*. Oxford University Press.
- Diamond, P. (1967). Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of Political Economy* 75(5), 765–766.
- Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics* 75(4), 643–669.

- Hansson, B. (1988). Risk aversion as a problem of conjoint measurement. In P. Gärdenfors and N.-E. Sahlin (Eds.), *Decision, Probability, and Utility*. Cambridge University Press.
- Harsanyi, J. C. (1977). On the rationale of the Bayesian approach: Comments on Professor Watkins's paper. In R. Butts and J. Hintikka (Eds.), *Foundational Problems in the Special Sciences*. D. Reidel Publishing.
- Jeffrey, R. (1965). *The Logic of Decision*. The University of Chicago Press.
- Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2), 263–292.
- Lewis, D. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability*. University of California Press.
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization* 3(5), 323–343.
- Savage, L. (1954). *The Foundations of Statistics*. John Wiley & Sons.
- Stefánsson, H. O. (2014). *Risk and Rationality* by Lara Buchak. *Economics and Philosophy* 30(2), 252–260.
- Stefánsson, H. O. (2015). Fair chance and modal consequentialism. *Economics and Philosophy* 31(3).
- Stefánsson, H. O. and R. Bradley (2015). How valuable are chances? *Philosophy of Science* 82(4), 602–625.
- Tversky, A. and P. Wakker (1995). Risk attitudes and decision weights. *Econometrica* 63(6), 1255–1280.
- von Neumann, J. and O. Morgenstern (1944). *Games and Economic Behavior*. Princeton University Press.
- Voorhoeve, A., K. Binmore, and L. Stewart (2012). How much ambiguity aversion? *Journal of Risk and Uncertainty* 45(3), 215–38.
- Wakker, P. (2010). *Prospect Theory: For Risk and Uncertainty*. Cambridge University Press.

Watkins, J. (1977). Towards a unified decision theory: A non-Bayesian approach. In R. Butts and J. Hintikka (Eds.), *Foundational Problems in the Special Sciences*. D. Reidel Publishing Company.