

Economics and Philosophy

<http://journals.cambridge.org/EAP>

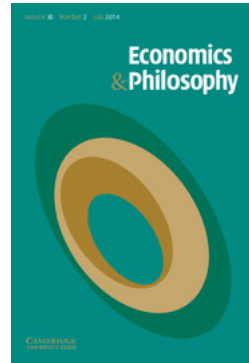
Additional services for ***Economics and Philosophy***:

Email alerts: [Click here](#)

Subscriptions: [Click here](#)

Commercial reprints: [Click here](#)

Terms of use : [Click here](#)



***Risk and Rationality*, Lara Buchak. Oxford University Press, 2013, xii + 256 pages.**

H. Orri Stefánsson

Economics and Philosophy / Volume 30 / Issue 02 / July 2014, pp 252 - 260
DOI: 10.1017/S0266267114000200, Published online: 02 June 2014

Link to this article: http://journals.cambridge.org/abstract_S0266267114000200

How to cite this article:

H. Orri Stefánsson (2014). *Economics and Philosophy*, 30, pp 252-260
doi:10.1017/S0266267114000200

Request Permissions : [Click here](#)

doi:[10.1017/S0266267114000200](https://doi.org/10.1017/S0266267114000200)

Risk and Rationality, Lara Buchak. Oxford University Press, 2013, xii + 256 pages.

Orthodox decision theorists hold that to determine what it is instrumentally rational for a person to do, we only need to know two things: how much the person desires different outcomes and her beliefs about how likely each of her alternatives¹ is to realize these outcomes. Similarly, according to the received picture in many of the social sciences (in particular economics), to predict the behaviour of a rational agent we only need to know how much she desires different outcomes and her beliefs about how likely her alternatives are in producing each outcome. Decision theorists and economists of course recognize the phenomenological reality of various attitudes beside desires and beliefs. But the idea is that any attitude that affects a person's behaviour can be modelled as either being part of the person's desires or beliefs.

Lara Buchak thinks that there is an important ingredient missing from this standard picture: people's willingness (or not) to accept the risk of something bad in exchange for a chance at something good. Attitudes to risk play a distinct role in rational action, she argues, and decision theorists should therefore distinguish these attitudes from both desires and beliefs. Two individuals might have the same desires and identical beliefs, but if one of them is prudent while the other is venturesome, then they might nevertheless differ in their preferences over alternatives. And neither of them needs to be irrational in the decision theoretic sense, she claims: both may have perfectly consistent attitudes and always prefer the means to their ends.

Buchak discusses various examples where attitudes towards risk influence the decisions people make. I will focus on two of the more familiar of her examples. The simplest example concerns a lottery that gives each participant a 0.5 chance of winning £100 but holds the same risk of yielding nothing. Ann, who is risk neutral, is willing to pay up to and including £50 to play in such a lottery. Bob, who is risk averse, is however only willing to pay up to £20 to take part in the lottery. Orthodox decision theory, also known as *Expected Utility Theory* (EU theory), models the difference between Ann and Bob as a difference in their desires over outcomes. Bob's desires are represented by a utility function that is concave with respect to monetary gains between £0 and £100; which, if we take the utility function to correctly represent Bob's desires, suggests

¹ To keep things simple, I will in what follows simply talk about *alternatives*, as the objects of preference, and not for instance distinguish between *acts* and *lotteries*.

that the desirabilistic difference between winning £50 and £0 is greater, according to him, than the difference between winning £100 and £50. Ann's desires are however represented by a *linear* utility function over money; suggesting that an extra dollar is equally valuable to her at all points in the interval in question.

But there is something dissatisfying about this solution, Buchak argues. Intuitively, the difference in Ann and Bob's attitudes need not have anything to do with how they value *outcomes*. Let's assume that Bob strongly opposes the claim that an extra pound is worth less to him the more money he wins: his unwillingness to evaluate the lottery in question according to its expected monetary value has nothing to do with how he values the possible prizes of the lottery, he claims, but simply how he trades-off the chance of something good against the risk of something bad.

Most decision theorists and economists will not be persuaded by the above argument (as Buchak acknowledges). According to what seems to be the received view amongst such theorists, a utility function is nothing but a numerical representation of an agent's preferences. Beyond that, there is no fact about an agent that corresponds to the utility function that represents her. Therefore, it is not a problem if Bob's phenomenological experience of his desires does not fit the modelling assumption that an extra pound is worth less to him the more money he wins: all that matters is that his preferences can be represented *as if* he were maximizing the expected value of a utility function with such a form.

Unfortunately for EU theorists, they cannot employ the same trick in another example of the impact of risk on people's choices: the infamous *Allais Paradox*. In the set up that generates paradox, people are asked to choose between two pairs of alternatives, or lotteries, firstly between L_1 and L_2 and secondly between L_3 and L_4 . Here is one way of presenting the alternatives, where, for instance, an agent who chooses alternative L_1 wins a prize of £5 million in the event that one of tickets 2–11 is drawn:

	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

When faced with this pair of alternatives, most people prefer L_1 to L_2 and L_4 to L_3 (a preference I will call the *Allais Preference*). The reasoning behind this preference, according to Buchak (12), is roughly the following. The minimum one stands to gain from L_1 and L_2 is the same, and the extra risk of winning nothing is only slightly increased when choosing alternative

L_1 , which gives a reasonable chance at a much higher prize than L_2 . So in that case, the extra risk of winning nothing seems to be worth the shot at the much higher prize. The minimum one stands to gain from L_4 is however a great deal higher than the minimum one stands to gain from L_3 . So in that case, the chance of a higher prize is not enough to offset the risk of ending up with nothing.

There is nothing *prima facie* irrational about the above reasoning, Buchak points out: this reasoning does not seem to have revealed any inconsistency in the reasoner's attitudes, who seems to be trying to identify the best means to her ends. But the problem is that no orthodox utility function can assign a higher value to L_1 than L_2 and also a higher value to L_4 than L_3 . So irrespective of how people with the Allais Preference value outcomes, they cannot (given this description of the alternatives) be represented as maximizing expected utility.² Hence, since utility theorists assume that a necessary condition for a person to be instrumentally rational is that her preferences can be represented as maximizing expected utility, they seem to be driven to the unpalatable conclusion that people with Allais Preference are actually irrational.

Buchak's *Risk-Weighted Expected Utility Theory* (REU theory) is specifically designed, first, to being able to capture the difference between Ann and Bob without modelling it as a difference in how they evaluate outcomes, and, second, to be able to represent intuitively rational preferences like Allais' as maximizing the expected value of a *risk-weighted* utility function. One of the main differences between REU theory and the more traditional EU theory concerns how many variables are determined by the agents we are trying to model. The orthodox theory leaves it up to the agent to decide, firstly, the values of potential outcomes, and, secondly, the probability of the different contingencies that affect whether these outcomes are realized. In addition, REU theory leaves it up to the agents to decide what strategy to use when aggregating the values of different possible outcomes of an alternative in order to evaluate the overall value

² I should point out that a popular response to the Allais Paradox is to argue that the above table does not correctly represent how most people view the four alternatives. In particular, winning nothing (£0) after having made the more risky choice in the second decision problem is worse, so the argument goes, than winning nothing after having made the more risky choice in the first decision problem, since in the second case but not the first it is true that had one chosen differently, one would *certainly* have won something. John Broome (1991) for instance responds to the paradox by adding 'disappointment' to the £0 outcome in the second decision problem. In 'Counterfactual Desirability' (working paper), Richard Bradley and I avoid the Allais Paradox by enriching the domain of decision theoretic value functions to *counterfactual* prospects. (The paper is a further exploration and application of the new decision theory I discuss at the end of Stefánsson, H. O. 'Fair chance and modal consequentialism', forthcoming in *Economics and Philosophy*.) Chapter 4 of *Risk and Rationality* contains various interesting arguments against these types of responses to the Allais Paradox.

of the alternative. The idea is that the strategy an agent uses will depend on how she trades-off chances for good outcomes against risks of bad outcomes. Whereas EU theory models rational agents as maximizing expected utility relative to a pair of utility and probability functions, REU theory models rational agents as maximizing expected utility relative to a triple of utility, probability and risk functions.

Risk and Rationality contains an exceptionally clear explanation of the formal difference between REU and EU theory, both in terms of how each theory measures the value of alternatives and the axiomatic difference between these measures. I will start by briefly summarizing the former difference. Let A, B , etc., be alternatives that will result in different outcomes, O_1, O_2 , etc., depending on which state, s_1, s_2 , etc., turns out to be actual. (Sets of states are called *events*.) Sometimes it will be convenient to use $A(s_i)$ to denote the outcome of alternative A if state s_i happens to be actual. Let u be a utility function defined over the set of outcomes and p a probability function defined over the set of states. To keep things simple, let us assume for now that alternative A will result in one of two possible outcomes, O_2 or O_1 , depending on whether event E actualizes or not (in the latter case I will say that event $\neg E$ actualizes). According to the traditional EU theory, the choice-worthiness of A is just its expected utility, as given by: $EU(A) = u(O_2) \cdot p(E) + u(O_1) \cdot p(\neg E)$. Since $p(\neg E) = 1 - p(E)$, the expected utility of A can be equivalently formulated as:

$$(1) \quad EU(A) = u(O_1) + p(E)[u(O_2) - u(O_1)]$$

The second way of formulating expected utility brings out very clearly the intuitive difference between EU theory and REU theory. The latter theory introduces a (non-decreasing) risk function, r , 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. Suppose that $u(O_1) < u(O_2)$. Then the choice-worthiness of an alternative according to REU, its *risk-weighted* expected utility, is given by:

$$(2) \quad REU(A) = u(O_1) + r(p(E))[u(O_2) - u(O_1)]$$

Since the better outcome O_2 occurs only if event E actualizes, the risk-weighted expected utility of A is given by the utility of its worse possible outcome, plus the *added* value if the better outcome occurs weighted by both the probability of the better outcome occurring (p) and how important the agent considers the better outcome relative to the worse outcome (r). If $r(p) = p$, then (2) is equivalent to (1), and the agent in question is an expected utility maximizer. If however $r(p) < p$, then the agent is risk averse, but if the opposite holds then the agent is risk-seeking.

Removing the simplifying assumption that A can only result in one of *two* possible outcomes, and assuming $u(O_1) \leq u(O_2) \dots \leq u(O_n)$, the REU of A is given by:

$$(3) \quad \text{REU}(A) = u(O_1) + \sum_{j=2}^n \left[r \left(\sum_{j=i}^n p(E_i) \right) (u(O_j) - u(O_{j-1})) \right]$$

As promised, the risk function allows us to represent the difference between Ann and Bob without modelling it as a difference in how they evaluate outcomes: whereas Ann has a linear risk function, Bob's is convex. Moreover, unlike EU theory, REU theory can, as Buchak shows, represent the Allais Preference as maximizing an expected value.

Equation (3) satisfies the natural condition that an alternative will never be worse than its worst possible outcome and never better than its best possible outcome. In addition, since r is non-decreasing, an agent whose preferences can be represented as maximizing REU will satisfy the familiar (Stochastic) *Dominance* principle, meaning that if for any outcome some alternative A has at least as high chance as alternative B for getting that or a better outcome, then the REU maximizer will never prefer B to A . Similarly, such an agent will always satisfy transitivity.

In fact, in terms of what decision theorists have traditionally wanted to count as *Rationality Axioms*, the only difference between a REU maximizer and an EU maximizer is that unlike the former, the latter satisfies what has become known as *The Sure Thing Principle* (STP).³ To formally state the principle, let $A_E x$ denote the alternative that agrees with alternative A on all states contained in E but otherwise yields outcome x ; let \mathbf{A} be the set of alternatives, \mathbf{O} the set of all possible outcomes of these alternatives, \mathbf{E} the set of possible events and \mathbf{S} the set of all possible states; and finally let \preceq denote the relation of preference-or-indifference (and $<$ strict preference). Here is then (one version of) the STP:

$$(4) \quad \text{For all } A, B \in \mathbf{A}, x, y \in \mathbf{O}, \text{ and } E \in \mathbf{E} : A_E x \preceq B_E x \Leftrightarrow A_E y \preceq B_E y.$$

Intuitively, the axiom says that if an agent would (at least weakly) prefer to replace the part of an alternative that occurs if E with B rather than A , then the same should hold *no matter what occurs if $\neg E$* . An implication of the axiom is that when comparing two alternatives like A and B you only need to consider the states of the world where A results in different

³ In her representation theorem for REU theory (which combines work done by Veronika Köbberling and Peter Wakker with work by David Schmeidler and Mark Machina), Buchak uses an axiom that is slightly different from the CSTP that I discuss below. But she points out that the (substantive) disagreement between EU and REU theorists can just as well be characterized as a disagreement over STP vs. CSTP. Since the STP has received a great deal of attention, in particular in the discussion of the Allais Paradox, I will focus on the difference between STP and CSTP.

outcomes than B . To see why someone with the Allais Preference violates STP (given how the alternatives are usually presented), notice that when we exclude the events where L_1 and L_2 differ, and do the same for L_3 and L_4 , then L_1 becomes identical to L_3 and L_2 to L_4 . Hence, a preference for L_1 over L_2 but L_4 over L_3 may seem to reveal some sort of inconsistency in the agent's attitudes.

Instead of satisfying the STP, a REU maximizer satisfies a principle called the *Comonotonic* Sure Thing Principle (CSTP). Again, Buchak's discussion of the difference between the STP and CSTP is exceptionally clear and illuminating, as is her explanation as to why an agent needs to satisfy something like either of these principles for it to be possible to represent her desires by a quasi-unique utility function. Before explaining CSTP, we need to define what it means for alternatives to be *comonotonic*.

Comonotonicity: Two alternatives, A and B , are said to be *comonotonic* just in case there are *no* states $s_i, s_j \in \mathbf{S}$, such that $A(s_i) \prec A(s_j)$ but $B(s_j) \prec B(s_i)$.

Less formally: two alternatives are comonotonic just in case they order possible states (and thus events), from 'good' to 'bad', in the same way. Here is a formal statement of CSTP:

- (5) For all $A, B \in \mathbf{A}$, $x, y \in \mathbf{O}$, and $E \in \mathbf{E}$: $A_E x \preceq B_E x \Leftrightarrow A_E y \preceq B_E y$
if all four alternatives are comonotonic.

It is easy to see that STP is strictly logically stronger than CSTP: anyone satisfying the former will satisfy the latter, but the converse does not hold. In the context of the other axioms of EU theory, the STP in effect guarantees that for any two alternatives, A and B , a particular outcome or sub-act contributes the same value towards the overall value of A as to the overall value of B . CSTP however only guarantees that this holds for alternatives that order the events in the same way. With this in mind it is easy to see that the Allais Preference does not violate CSTP: the event that ticket 1 is drawn is strictly worse than any other event when alternative L_3 has been chosen, but the same is not true given any of the other alternatives. So the alternatives are not comonotonic.

Buchak makes various arguments intended to show that an agent who satisfies CSTP but not STP need not be instrumentally irrational: her attitudes need not be inconsistent and she may prefer the means to her ends. Buchak for instance shows that arguments from dominance reasoning do not support the claim that agents who violate STP but not CSTP fail to be rational (Chapter 5); argues that it need not be irrational to prefer less information to more, as those who violate STP but satisfy CSTP will sometimes do (Chapter 6); and points out that neither *Dutch Books* nor arguments that appeal to the long run show that EU maximizers are more instrumentally rational than REU maximizers (Chapter 7).

Buchak should be applauded for illustrating and confronting, rather than trying to hide, what may seem to be weaknesses of REU theory (e.g. the implication that rational agents may prefer not to get what may seem to be perfectly good information). And her defence of the theory against potential criticism is for most parts convincing. Nevertheless, I would, before concluding, like to mention one thing that I find to be a weakness of *Risk and Rationality*. The issue concerns how to interpret the risk function. It has become commonplace in philosophy to distinguish beliefs from desire by their *direction of fit*: beliefs are those attitudes we try to fit to the world (i.e. we change the attitude rather than the world when the two do not fit), whereas desires are those attitudes we try to fit the world to (i.e. we change the world rather than the attitude when there is a misfit). The risk function represents neither of these types of attitudes, Buchak says:

[I]t 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 to desires. Instead, it measures how an agent *structures the potential realization of some of his aims*. (53–54; emphases added)

Given that REU theory is a normative theory, it had better not be the case that attitudes to risk are part of an agent's *beliefs*. For that would e.g. imply that a risk-averse person believes that a particular event *E* becomes more likely when an alternative is chosen that results in a bad outcome if *E* occurs. On pp. 110–114 Buchak convincingly argues that it is most natural to interpret *r* as measuring something distinct from an agent's beliefs. But I am not quite convinced that we should not take the risk function to be a measure of (a part of) an agent's *desires*.

Buchak describes the risk-averse Jeff, who (like other risk-averse people) 'has a preference about how values are arranged across the possibility space' (29). Rather than concentrating good outcomes in one part of the possibility space and bad outcomes in another, which would mean that if he were lucky (unlucky) concerning what state is actual he would get something *very* good (bad), he prefers that good and bad outcomes are distributed evenly over the possibility space such that, if possible, he is guaranteed to get at least something that is not *too* bad even if he is unlucky about what state is actual. And I take it that if rational, Jeff will, whenever possible, make choices that correspond to this preference: he will try to make sure that the alternatives – or 'gambles' – he holds distribute good outcomes as evenly as possible across the possibility space.

To me, this sounds very much like risk attitudes are part of an agent's desires. Risk-averse agents *desire* that outcomes be arranged one way over the possibility space, risk-loving agents another way, and when possible, they change an important part of the world to better fit these attitudes; namely the gambles they hold. The representation theorem Buchak provides for REU theory of course shows that we can represent a particular class of agents as maximizing expected utility with respect to a probability, utility *and* risk function, where the latter two each play a distinct role. But that in itself does not establish that the utility and risk functions do not represent for instance two different aspects of an agent's desires.

Buchak emphasizes in various places that she takes the risk function to represent how an individual 'structures the potential realization of some of his aims'. But again, it is unclear to me that this is incompatible with seeing the risk function as representing part of an agent's desires. Suppose a risk averse agent can realize some particular aim α in one of two ways, by choosing either A or B , both of which she expects to realize α to degree x (and neither of which can either positively or negatively contribute to the realization of any other of her aims). Then the agent will, I would suggest, *desire* more strongly the alternative that has the potential to realize α in a less risky way (in other words, the one that distributes the degrees to which α is realized more evenly across the space of possible outcomes); thus *desiring* to structure the potential realization of α in as risk-free way as possible.

Is it really important, one might ask, whether or not we interpret r as representing part of an agent's desires? One reason the interpretation matters, is that if r represents parts of an agent's desires, then we might want to know more about the relationship between the desires that this function represents and the agent's desires for specific outcomes. Another reason why the interpretation might matter, is that if risk attitudes truly are part of desires, then the need for *risk-weighted* expected utility theory is much less well motivated. Expected utility theorists typically respond to the Allais Paradox by suggesting that we re-describe the relevant alternatives (in particular, their consequences) in a way that makes it possible to represent the Allais Preference as maximizing the expected value of an orthodox utility function, typically by adding some variable \hat{d} to the zero-outcome in alternative L_3 .⁴ In Chapter 4, Buchak shows that the most straightforward justifications and strategies for such re-descriptions face certain problems. Nevertheless, if risk aversion is part of an agent's desires, then perhaps *some* such re-description is in order. Buchak herself points out that if, for instance, an agent desires £100 that she has earned more strongly than £100 that she has stolen, then to model

⁴ See for instance Broome (1991: ch. 5).

that person's attitudes we need to distinguish between outcomes where the agent receives £100 in these different ways. Similarly, if attitudes to risk are a part of our desires, then when modelling agents who are not risk-neutral, we should perhaps always distinguish the *outcome* of winning £100 as a result of a risky alternative from the outcome of winning £100 as a result of a risk-free alternative.

In spite of disagreeing with Buchak's interpretation of the risk function, I would like to end this review on a positive note. Given the technical nature of its subject matter, *Risk and Rationality* is surprisingly easy to follow. Buchak's writing is free of jargon and she manages to explain even the technical details of her theory in such a non-technical way that I am sure any student of philosophy will be able to follow her discussion. The book moreover contains very interesting passages on what we might call 'the philosophy of decision theory', such as metaphysical and epistemological issues concerning utilities and probabilities. Her discussion of the main differences between her new decision theory and the orthodoxy is so clear and illuminating that even those who remain unconvinced of the need for a new normative theory of rational choice should, after having read *Risk and Rationality*, be in a better situation to state exactly why they prefer the orthodoxy.

H. Orri Stefánsson

London School of Economics and Political Science, UK

REFERENCE

Broome, J. 1991. *Weighing Goods*. Oxford: Blackwell Publishers.