

Notre Dame Philosophical Reviews

2015.03.14

Author ▼

Lara Buchak

Risk and Rationality

Published: March 09, 2015

Lara Buchak, *Risk and Rationality*, Oxford University Press, 2013, 260pp., \$60.00 (hbk), ISBN 9780199672165.

Reviewed by Rachael Briggs, Australian National University

Among decision theorists, expected utility theory (EU theory henceforth) is the reigning orthodoxy. EU theory aims to give an account of ideal means-end rationality; it provides constraints that agents must satisfy to pursue their ends in a consistent way. This book challenges one of EU theory's central commitments: that ideal rationality is *risk-neutral*. Lara Buchak develops an alternative to EU theory that lets decision theorists model positive or negative attitudes toward risk. She presents her theory as a rival account of ideal rationality, rather than a descriptive account of human choice, or a complementary account of non-ideal rationality.

EU theory says that an agent must choose the act with the highest expected utility. Where her act has a finite number n of possible outcomes (which we can number from the worst, x_1 , to the best, x_n); P_i is the probability of getting an outcome at least as good as x_i on the supposition that act A is performed; $u(x_i)$ is the utility that the agent assigns to outcome x_i ; and $u(x_0) = 0$, this means the agent must choose the act A that maximizes:

$$EU(A) = [\text{sum}] P_j (u(x_j) - u(x_{j-1})) \text{ [with the sum ranging from 1 to } n\text{].}$$

The idea is that to calculate the expected utility of A , you take the utility of its worst possible outcome x_1 as a lower bound, increase that by the amount of additional utility provided by its next-worst possible outcome x_2 , weighted by the probability of getting an outcome at least as good as x_2 , add the amount of extra utility beyond *that* provided by its third-worst possible outcome x_3 , weighted by the probability of getting an outcome at least as good as x_3 , and so on.

EU theory is best understood as a set of necessary conditions on rational choice: it tells agents how to be consistent, but not what substantive ends they should aim toward, or what beliefs should guide their deliberations. Buchak claims that EU theory's requirements, minimal as they are, are still too strong; a person can violate them without being irrational.

The problem with EU theory, according to Buchak, is that it takes only *local* properties of gambles into account. If a gamble comes with a 50% chance of winning a million dollars, then that chance makes the same contribution to the overall value of the gamble regardless of the other possible outcomes and their probabilities. Buchak, on the other hand, thinks that the value contributions of outcomes are context-dependent, and influenced by global properties of gambles. For a risk-averse person, a 20% chance of winning a million dollars may count more than twice as much toward the value of a lottery as a 10% chance of winning a million dollars, even though the value of the million-dollar prize is independent of how it was obtained.

Buchak illustrates the failings of EU theory with a collection of examples, including the famed Allais paradox. The Allais paradox involves four betting arrangements constructed from three prizes (Bad, OK, and Good) and three events (Probable, Improbable, and Very Improbable). In the matrix below, each row corresponds to a betting arrangement, while each column corresponds to an event. The cells in the matrix list the payoff yielded by the betting arrangement, in the event of the column.

Probable (89%)

Improbable (10%)

Very Improbable (1%)

L1 OK with certainty	OK	OK	OK
L2 1% Chance of Bad; 10% chance of Good; OK otherwise	OK	Good	Bad
L3 11% chance of OK; Bad otherwise	Bad	OK	OK
L4 10% chance of Good; Bad otherwise	Bad	Good	Bad

A reasonable person might prefer the safe L1 to the risky L2, and at the same time, prefer L4, with its slightly lower chance of the Good prize, to L3, with its slightly higher chance of the merely OK prize. But there is no way of assigning utilities to the prizes Bad, OK, and Good that makes these preferences consistent with EU theory.

There are three ways to respond to the Allais paradox (and the related puzzle cases Buchak considers): reject EU theory, claim that the apparently reasonable preferences are really irrational, or re-describe the choice situation in a way that makes the preferences consistent with EU theory. Buchak rejects the second and third solutions, and defends the first. But if we needn't obey EU theory, what must we do? Buchak answers this question by developing an alternative to EU theory, which she calls risk-weighted expected utility theory, or REU theory.

According to REU theory, a rational agent should choose the act A that maximizes

$$REU(A) = [\text{sum}] r(P_i)(u(x_i) - u(x_{i-1})) \text{ [with the sum ranging from 1 to } n].$$

This is like the equation for EU theory, but with a twist: instead of weighting the extra value contributed by each outcome by the probability that it (or a better outcome) will occur, the agent weights the value of each outcome by a function r of the probability that it (or a better outcome) will occur. If r is concave, then the agent assigns less weight than an EU theorist to chancy outcomes whose utility exceeds the minimum; she is risk-averse. If r is convex, then the agent assigns more weight than an EU theorist to chancy outcomes whose utility exceeds the minimum: she is risk-seeking. Other, more complicated attitudes toward risk are possible, since risk functions can have both convex and concave segments.

For EU theorists, the instrumental value of an act depends on two parameters: a function that assigns utilities to its possible outcomes, and a function that assigns probabilities to those outcomes, conditional on each act being chosen. Buchak adds a third parameter: a continuous, increasing function r , which maps the $[0, 1]$ interval to itself. r measures the agent's attitude to risk -- a global property of entire gambles, rather than a local property of individual outcomes. Like probability and utility functions, r can vary from agent to agent. REU theory is thus a generalization of EU theory: EU theory is the special case where $r(p) = p$, for all p .

The first half of the book develops REU theory in detail. Buchak carefully distinguishes between formal representations and their philosophical interpretation. (Particularly helpful are the discussion of formalist, constructive realist, and non-constructive realist interpretations of credence and utility in section 1.2, and the discussion of interpretive vs. prescriptive understandings of decision theory in section 3.1.) This part of the book culminates in a representation theorem that links unobservable probabilities, risk functions, and utilities to more easily observable preferences over acts.

The theorem roughly follows the technique of Leonard Savage, postulating three separate but linked domains of events (the bearers of probability), outcomes (bearers of utility), and acts (bearers of risk-weighted expected utility). Buchak gives a set of constraints on rational preferences, and shows that the agent who satisfies them is representable by a unique probability function, a unique risk function, and a utility function that is unique up to positive linear transformation.

The representation theorem inherits the weaknesses of the Savage framework: it fails to account for probabilistic dependence between acts and states, it requires an implausibly rich space of acts, it can only represent credences that match the agent's betting odds, and it has trouble translating 'grand-world' problem descriptions into manageable 'small-world' problem descriptions.

But these weaknesses are not exclusive to Buchak's view; they are shared by any version of EU theory that relies on a Savage-style representation. Nor are these weaknesses essential to Buchak's view. EU theory has alternative representation theorems, such as those championed by Frank Ramsey and Richard Jeffrey (though these representation theorems face problems of their own). Rather than being a damning objection, this criticism points to a promising research question: how might alternative representation theorems be adapted to characterize REU theory without building in Savage's problematic assumptions?

In the second half of the book, after presenting the representation theorem, Buchak addresses objections to REU theory. All but one of these objections are linked to the charge that risk-seeking and risk-aversion are forms of inconsistency or means-end irrationality. (The exception is in Chapter 4, which argues that we should not re-describe the choices in the Allais paradox in a way that makes the contentious preferences compatible with EU theory.)

A root cause of REU theory's alleged inconsistency is that it allows violations of the

Sure-Thing Principle: If A and B are two acts that yield the same outcomes conditional on E , then the agent's unconditional preference between A and B matches the agent's conditional preference between A and B given not- E .^[1]

We can construct scenarios where a risk-averse (or risk-seeking) agent strictly prefers A to B , but strictly prefers B to A both conditional on E and conditional on not- E , in violation of the Sure-Thing Principle.

Here is an illustrative example. I can either stick to the status quo (utility = 0) with certainty, or pledge to volunteer at my local animal shelter (in which case a friend will donate to the shelter in my honor). The animal shelter is unlikely to need volunteers; there is an 89% chance that I will never be called upon, the donation will improve on the status quo (utility = 1). There is a 10% probability that I will be called upon to volunteer, and will have an excellent experience (utility = 10). And there is a 1% probability that I will be called upon to volunteer and be bitten by an animal (utility = -10). If my risk function is $r(p) = p^2$, then I will think pledging to volunteer is too risky; the risk of animal attack outweighs the value of the improbable good experience. But conditional on my being asked to volunteer, the good experience becomes likely enough to outweigh the risk of attack (probability 10/11, or a little under 91%), so I would rather volunteer than not.

Chapter 5 rebuts the objection that the Sure-Thing Principle is a self-evident principle of rationality. Buchak helpfully distinguishes the Sure-Thing Principle from several weaker principles that follow from REU theory, including a statewise dominance principle. While these weaker principles are principles of rationality, none of them entails the Sure-Thing Principle.

Chapters 6 and 7 rebut the objection that REU theory sometimes advises agents to choose dominated courses of action. Chapter 6 discusses two diachronic arguments, in which a clever bookie persuades the risk-averse agent to choose one sequence of acts, even though there is an available sequence of acts that yields a better outcome in every state of the world. Both arguments involve a scenario where the agent prefers A to B , but her preference is reversed both conditional on E and conditional on not- E .

In the first argument, the bookie offers the agent a choice between A and B , tells her whether E , and then makes her pay to switch -- which she will be willing to do whether E is true or false. In the second argument, the bookie offers the agent a choice between A and B , and makes her pay not to learn E (and so not to reverse her preference) until after she has made the choice.

Buchak argues that the bookie's two strategies rely on two different and incompatible assumptions about the relationship between rational choice and preference. The pay-to-switch strategy assumes a *naïve* model of choice, in which the agent always chooses the next step in the sequence of steps that she currently prefers best. The pay-to-avoid-information strategy assumes a *sophisticated* model of choice, in which an agent always chooses the best available step given what she knows about her future

choices. There is also a *resolute* model, in which the agent settles on a preferred course of action at the beginning of the decision situation, and sticks to her plan at all subsequent choices. Agents who use the resolute models can never be exploited by clever bookies.

Buchak argues that the naïve model is a poor one, and that REU theorists can (and should) adopt either the sophisticated or the resolute model. This seems to leave REU theorists with a problem: if the sophisticated model is the right one, then doesn't REU theory advise a dominated sequence of acts in the pay-to-avoid-information scenario? Buchak argues that it does not. True, sophisticated choosers will pay to avoid information. But the dominating sequence -- refusing to pay, then choosing *A* over *B* anyway -- is not really available. Although it is up to the agent whether to pay, and up to the agent whether to choose *A* over *B*, sequences of actions are not direct objects of choice at all.

The dominance arguments in Chapter 7 are synchronic, involving multiple bets made at the same time. A risk-seeking agent will be willing to bet on worse than 50-50 odds on a proposition and on its negation -- but when both these bets are taken together, they result in a sure loss. Buchak argues that while REU theory does tell the risk-seeking agent to take each bets, if the value of the status quo is held fixed no matter which event occurs, it does not tell her to take both bets together. On REU theory, the values of bets are context-dependent -- they can vary according to whether the bet is bought or sold, and according to how risky the status quo is. Betting on *E* when one has already bought a bet on not-*E* is different from betting on *E* when one has no outstanding bets.

Should we reject EU theory in favor of REU theory? There are costs and benefits on both sides. On the one hand, REU theory can explain and accommodate preferences that are hard for EU theorists to rationalize. On the other hand, REU theory might rationalize too much: it endorses sets of acts that, together, form dominated sequences (at least if the sophisticated model of choice is correct). And REU theory makes it hard to factor the values of uncertain prospects into the values of context-independent sub-lotteries or bets.

Perhaps we should side with Buchak, and hold that the balance of evidence favors REU theory. Another possible resolution of the problem (not endorsed by Buchak) is to view REU theory as a supplement, rather than a rival, to EU theory. Perhaps there is no single, unified ideal of rationality, but a variety of levels of idealization that are useful in different circumstances.

Regardless of the meta-normative facts about rationality, *Risk and Rationality* is a valuable addition to any decision theory library. Buchak provides an exceptionally clear, careful, and sustained defense of a theory for reasoning about uncertainty that takes risk into account.

[1] My phrasing of the Sure-Thing principle differs from Buchak's in particulars, but not in its essence. For ease of exposition, I've framed the principle in terms of conditional preference, while Buchak's version is framed in terms of unconditional preference. Also Buchak's version is restricted to cases where *A* and *B* yield the same outcome in every state compatible with not-*E*.