The Economics and Philosophy of Risk

H. Orri Stefánsson
Stockholm University
and Swedish Collegium for Advanced Study
orri.stefansson@philosophy.su.se
www.orristefansson.is

Abstract

Neoclassical economists use expected utility theory to explain, predict, and prescribe choices under risk, that is, choices where the decision-maker knows—or at least deems suitable to act as if she knew—the relevant probabilities. Expected utility theory has been subject to both empirical and conceptual criticism. This chapter reviews expected utility theory and the main criticism it has faced. It ends with a brief discussion of *subjective* expected utility theory, which is the theory neoclassical economists use to explain, predict, and prescribe choices under uncertainty, that is, choices where the decision-maker cannot act on the basis of objective probabilities but must instead consult her own subjective probabilities.

Keywords: *Risk; Risk Aversion; Uncertainty Neoclassical Economics; Expected Utility; Rationality.*

1 Introduction

Decisions often have to be made without knowing for sure which outcome will result from her choice. When deciding between taking the train and driving to work, for instance, one may be unsure about several factors that could determine the outcome of these choices, such as whether the train will be on time and whether an accident will cause congestion on the road. Similarly, important economic decisions, such as choosing between pension-plans, have to be made with imperfect knowledge of crucial factors, for instance, how long one will live and the actual returns of the different pension-plans. A common way to put this, is that decisions such as these are not made in situations of *certainty*.

Decisions that are not made in situations of certainty however differ widely in terms of how much or little the decision-maker knows. The extent of a decision-makers knowledge, in a particular situation, can be usefully characterised with reference to the elements of the decision-maker's *decision-problem*. A decision-problem, as I shall be using the term, consists of two or more *options*, a set of *outcomes* that each of these options could result in, and a set of *states of the world* (or simply states), that determine which outcome results from each option. In the above transport decision-problem, for instance, the options include driving and taking the train, the states include facts about how well the trains run and how much traffic there is on the road, and the outcomes include arriving on time and arriving late.

If the decision-maker is fortunate enough to know which state obtains, then her decision is made in a situation of certainty. When she does not know which state obtains, we can nevertheless draw important distinctions depending on how much the decision-maker knows about the states of the world. Sometimes, for instance, a decision-maker knows—or at least "deems suitable to act as if" she knew (Luce and Raiffa 1989/1957: 277)—the *probabilities* of the different states of the world. A game of roulette would typically be treated as a situation where the gambler knows the probabilities of the relevant states, and, by implication, the probabilities of the possible outcomes of choosing each option. For instance, a player of a European roulette knows that there is a probability of 1/37 that the ball ends up in pocket numbered 5 ("state"); hence, she knows that if she chooses to bet on number 5 ("option") then there is a 1/37 probability that she wins ("outcome"). Following Knight (1921), economic theorists typically use the term "risk", for such choices; and say that the roulette gambler is making a decision *under risk*.

In contrast, when betting on a soccer match, one does not know all the relevant probabilities. The outcome of the match may, for instance, depend on whether the hot-headed midfielder of the home-team gets a red card, whether the star-striker of the away-team gets injured, and so on. But the probability that the midfielder gets a red card is hard to know. We might know that he has so far received a red card in 10% of all games she has played, but we do not know whether her mood on the day in question will be better or worse than normal. And it might be even less plausible to say that we can know the probability with which the star-striker gets injured on the day in question. We might, however, know all the states of the world that could determine which of the (say, three) outcomes (home-team wins, away team wins, draw) obtains. In that case, economic theorists, again following Knight (1921), would use the term "uncertainty"; and they would say that the bettor is making a decision *under uncertainty* (but not under risk).

In some decision-problems a decision-maker's lack of knowledge is more severe than in either of the above two examples. In particular, sometimes a decision-maker might not even know all of the states of the world that could determine the outcome of her decision, and/or she might not be aware of all the possible outcomes that could result from her decisions. When evaluating the option of implementing solar geo-engineering as a response to the climate crisis, for instance, there plausibly are important states of the world and potential outcomes that we have not yet considered. More generally, when considering new technologies and radical policies we may suspect—e.g. based on past experience of similar decisions—that there are important contingencies that we cannot yet articulate. In addition, one might often not even know about all the available options; for instance, there presumably are some options for responding to the climate crisis that nobody has yet considered. Decisions where an agent lacks knowledge of some of the possible states, outcomes, ore options, are said to be made in situations of "unawareness" (for recent overviews of this literature, see e.g. Schipper 2015 and Steele & Stefánsson 2021).

This chapter will be almost exclusively concerned with how economic theorists (in particular, so-called "neoclassical" or "orthodox" economic theorists) treat risk, as previously defined. I shall start, in the next section, by outlining the theory that neoclassical economists use to predict, explain, and guide choices in situations of risk. In section 3 I then discuss some of the main challenges, both empirical and philosophical, to this orthodox treatment of risk. In section 4, I however briefly discuss how neoclassical economists tend to approach decision-making under uncertainty, and an important challenge faced by this approach. Section 5 concludes the chapter.

2 Risk in Economic Theory

Recall that a decision under risk is one where the relevant decision-maker knows, or acts as if she knew, the probabilities with which the available options deliver the possible outcomes. This is often described as decision-making with *objective* probabilities. But that terminology may be misleading. For instance, assuming that the behaviour of roulette wheels is deterministic, the arguably most common-sensical account of objective probabilities[1] would entail that the ball has a probability of 0 of ending up in any pocket except one, namely, the one that it will actually end up in, for which the probability is 1.

Nevertheless, betting on a roulette wheel is an archetypical example of decision-making under risk in economic theory, where the decision-maker is modelled as acting on the basis of knowledge of a non-trivial probability distribution, where "non-trivial" means that more than one outcome is assigned a positive probability. Hence, I shall use the term "known probabilities" rather than "objective probabilities" when describing decision-making under risk.

Now, my preferred terminology might be misleading too, since some might not find it appropriate to describe the roulette gambler as "knowing" the probabilities of the various outcomes. For instance, the epistemic sceptic might complain that given my terminology, nobody ever makes decisions under risk; while the determinist could argue that since one can only know that which is true, decisions under risk only involve trivial probabilities. So, the reader should keep in mind that when I speak of known probabilities, what I really mean is that the decision-maker finds it suitable to act as if she knew the relevant probabilities (Luce & Raiffa 1989/1957, ibid.).

It it also worth noting that a decision-maker may quite reasonably find it suitable to act as if she knew the relevant probabilities even when these are in fact not knowable. For instance, suppose that a patient is considering undergoing a surgery and learns that one million patients "like here", in the physiologically relevant sense, have undergone the surgery, and that the surgery has been successful in 99% of these cases. Or, suppose instead that a person is considering investing in government bonds and finds that all experts agree that there is at least a 90% chance that the bonds will yield a return of at least 5%. In these cases the *true* probabilities in question may not be knowable—for instance, perhaps the true probability with which a particular patient will have a successful surgery cannot be known. Nevertheless, it would seem reasonable for these decision-makers to act *as if they knew* the relevant probabilities: in the first case, that there is a 99% chance that the surgery will be a success; in the second case, that there is at least a 90% chance that the bonds will yield a return of at least 5%. So, I will treat examples like these as decision-making under risk.

The orthodox (neoclassical) approach in economics when it comes to explaining, predicting and guiding decision-making under risk is a theory that is often called *objec-*

[1]However, for a sophisticated account of objective probabilities that does not entail this, see Hoefer (2007) and Frigg & Hoefer (2010).

tive expected utility theory. "Objective" here refers to the probabilities assumed by the theory, rather than the utility. In what follows I shall simply call the theory in question "expected utility theory"; but later I shall consider *subjective* expected utility theory (where the probabilities are subjective).

Informally, expected utility theory says that the value of a risky option equals the option's *expectation* of utility, where "utility" is a measure of the desirability of the option's potential outcomes, and the expectation is calculated by multiplying each outcome's utility with its probability, and then adding up all of these probability-weighted utilities. To state this more precisely we need to introduce some formal definitions and notation.

2.1 The vNM theory

Let L_i be a "lottery" from the set $\mathbf{L}$ of lotteries, and O_k the outcome, or "prize", of lottery L_i that arises with probability p_{ik} (where, of course, $\sum_j p_{ij} = 1$). It is important to stress that the term "lottery" is a technical one; informally, it can be any risky option, that is, an option that could result in different outcomes, for which the decision-maker of interest knows (or acts as if she knows) the probabilities. The representation result that I discuss below requires the set $\mathbf{L}$ of lotteries to be rather extensive: it is closed under "probability mixture", which means that if $L_i, L_j \in \mathbf{L}$, then compound lotteries that have L_i and L_j as possible "outcomes" are also in $\mathbf{L}$. The expected utility of L_i is defined as:

vNM expected utility equation. $EU(L_i) = \sum_k u(O_k) \cdot p_{ik}$

According to expected utility theory, a rational preference between lotteries corresponds to the lottery's expected utilities, in the sense that the one lottery is preferred over another just in case the one offers a higher expectation of utility than the other. When this relationship between preference and expected utility holds, we say that the preference can be *represented as maximising expected utility*. (Why only "represented as"? Because the utility is simply a way of numerically describing the preference; no claim is made about utility corresponding to anything that the agent recognises. We shall get back to this issue soon.)

To state more formally the aforementioned relationship between rational preference and expected utility, we need some additional notiation. Let $\precsim$ denote a *weak* preference relation. So $A \precsim B$ means that the agent we are interested in considers option B to be at least as preferable as option A. From the weak preference relation we can define the *strict* preference relation, $\prec$, as follows: $A \prec B =_{\text{def}} A \precsim B \ \& \ \neg(B \precsim A)$, where $\neg X$ means "it is not the case that X". So, $A \prec B$ means that the agent prefers B to A. Finally, indifference, $\sim$, is defined as: $A \sim B =_{\text{def}} A \precsim B \ \& \ B \precsim A$. This means that the agent we are interested in considers A and B to be equally preferable.

Economists and decision theorists generally take there to be a close conceptual connection between preference and choice. Least controversially, a rational person who prefers B to A has a tendency to choose B over A, if given the option. More controversially, some economists have wanted to *define* preference (or "revealed preference"; Samuelson 1938, 1948) in terms of choice; to prefer B over A *means* having a tendency to choose B over A. How closely to tie preference to choice is an issue that we will have reasons to revisit.

We say that there is an expected utility function that represents the agent's preference $\precsim$ between lotteries in $\mathbf{L}$ just in case there is a utility function u and a probability function p such that for any$L_i, L_j \in \mathbf{L}$:

$$L_i \precsim L_j \Leftrightarrow EU(L_i) = \sum_k u(O_k) \cdot p_{ik} \leq EU(L_j) = \sum_k u(O_k) \cdot p_{jk}$$

Economist are, at least by tradition, skeptical of claims about people's attitudes that cannot, in principle at least, be reframed as claims about choice-behaviour.[2] Fortunately, claims about preferences can, at least in theory, be reframed as claims about (hypothetical) choice-behaviour, assuming that people generally (or at least rationally) choose what they prefer. Hence, there is no wonder that economists were impressed when von Neumann and Morgenstern (vNM) demonstrated that claims about utilities—for instance, the claim that a person maximises expected utility—can be reformulated as claims about a person's preferences between lotteries.[3] In particular, vNM proved that if a person's preferences between lotteries satisfy a number of constraints, or *ax-*

[2] Robbins (1932), Samuelson (1938, 1948), and Friedman (1953) are some influential works in this behaviourist tradition; for an overview, see Angner & Loewenstein (2012), in particular section 2.2.

[3] Ramsey (1990/1926) had actually already suggested a stronger result, that is, one that simultaneously derives a probability function and a utility function from the agent's preference (a project later continued by e.g. Savage 1972/1954 and Jeffrey 1990/1965). Nevertheless, this result of Ramsey's was never nearly as influential in economics as vNM's, perhaps partly since Ramsey neither gave a full proof of his result nor provided much detail of how it would go, but probably also partly since Ramsey's construction assumes certain psychological facts about agents (in particular, which prospects are consider "ethically neutral", that is, neither of negative nor positive value) that are prior to the expected utility representation (for a discussion, see Bradley 2001).

ioms, then she can be represented as maximising expected utility.

The following notation will be used to introduce the vNM axioms[4] of preference: $\{pA, (1-p)B\}$ denotes a lottery that results either in A, with probability p, or B, with probability $1-p$. $p \in [0,1]$ means that p takes a value between 0 and 1 (inclusive) whereas $p \in (0,1)$ means that p takes a value strictly between 0 and 1 (so, excluding 0 and 1). Note that the set the set **L** of lotteries contains "trivial" lotteries—that is, lotteries with only trivial probabilities—in addition to non-trivial ones. Since the "expected" utility of a trivial lottery equals, by the expected utility equation, the utility of the only outcome that it might result in, it follows, from a theorem we are about to state, that since these axioms hold for any lottery, they also hold for any outcome. The set of all possible outcomes is denoted **O**.

Axiom 1 (Completeness)**.** *For any $L_i, L_j \in \boldsymbol{L}$, either $L_i \precsim L_j$ or $L_j \precsim L_i$.*

Axiom 2 (Transitivity)**.** *For any $L_i, L_j, L_k \in \boldsymbol{L}$, if $L_i \precsim L_j$ and $L_j \precsim L_k$ then $L_i \precsim L_k$.*

Axiom 3 (Continuity)**.** *For any $L_i, L_j, L_k \in \boldsymbol{L}$, if $L_i \prec L_j \prec L_k$ then there is a $p \in (0,1)$ such that:*

$$\{pL_i, (1-p)L_k\} \sim L_j$$

Axiom 4 (Independence)**.** *For any $L_i, L_j \in \boldsymbol{L}$, if $L_i \precsim L_j$ then for any $L_k \in \boldsymbol{L}$, and any $p \in [0,1]$:*

$$\{pL_i, (1-p)L_k\} \precsim \{pL_j, (1-p)L_k\}$$

Axiom 5 (Reduction of compound lotteries)**.** *For any $L_i, L_j \in \boldsymbol{L}$, if for any $O_k \in \boldsymbol{O}$, $p_{ik} = p_{jk}$, then $L_i \sim L_j$.*

The Completeness axiom says that an agent can compare, in terms of the weak preference relation, all pairs of options (i.e., lotteries) in **L** and, by implication, all outcomes in **O**. Whether or not Completeness is a plausible rationality constraint depends, for instance, on what sort of options are under consideration, and how we interpret preferences over these options. If **O** includes all kinds of outcomes—e.g. curing cancer and eradicating poverty—then Completeness is not immediately compelling. If, on the other hand, all options in the set are quite similar to each other, say, all options are pension plans, then Completeness is more compelling.

[4]The axioms I present are not exactly the one's vNM presented. In fact, my choice of axioms is determined mainly by pedagogical reasons.

The plausibility of Completeness also depends on how closely we tie the interpretation of preference to *actual* choices, that is, choices that a person is actually faced with. As Gilboa (2009) notes, after having defined Completeness as a property of a weak preference relation:

> If we take a descriptive interpretation, completeness is almost a matter of definition: the choice that we observe is defined to be the preferred one. ... Taking a normative interpretation, the completeness axiom is quite compelling. It suggests that a certain choice has to be made. (51-52)

Here Gilboa is clearly thinking of preference in relation to decisions that the agent actually faces. As previously mentioned, however, the domain of for instance the preference relation in the vNM theory is far from containing only decisions that a person will actually face. Due to this, Aumann (1962) claimed that:

> Of all the axioms of utility theory, the completeness axiom is perhaps the most questionable. Like others of the axioms, it is inaccurate as a description of real life; but unlike them, we find it hard to accept even from the normative viewpoint. Does "rationality" demand that an individual make definite preference comparisons between *all* possible lotteries [...]? For example, certain decisions that our individual is asked to make might involve highly hypothetical stations, which he will never face in real life; he might feel that he cannot reach an "honest" decision in such cases. (446)

Few people would however question the plausibility of Transitivity as a requirement of rationality.[5] Informally, Transitivity says that if one option is at least as preferable as another option which is at least as preferable as a third option, then the first option is at least as preferable as the third option. To see why preference must be transitive for it to be possible to numerically represent it, it suffices to notice that if the first option gets assigned at least as high a number as the second option which gets assigned at least as high a number as the third option, then, necessarily, the first option gets assigned at least as high a number as the third option.

There is a straightforward defence of Transitivity that hinges on the sure losses that may befall anyone who violates the axiom (Davidson et al. 1955). This is the so-called

[5] But of course, for almost any claim, one can find a philosopher arguing against it. Notable philosophers who question the claim that Transitivity is a requirement of rationality include Temkin (1987, 1996, 2012) and Rachels (1998)

money pump argument. It is based on the assumption that if you find one option at least as preferable as another, then you should be happy to trade the one for the other. Suppose you violate Transitivity; for you: $L_i \precsim L_j$, $L_j \precsim L_k$ but $L_k \prec L_i$.[6] Moreover, suppose you presently have L_i. Then you should be willing to trade L_i for L_j. The same goes for L_j and L_k: you should be willing to trade L_j for L_k. You strictly prefer L_i to L_k, so you should be willing to trade in L_k plus some sum £x for L_i. But now you are in the same situation as you started, having L_i but neither L_j nor L_k, except that you have lost £x! This process could be repeated, the argument goes, thus turning you into a "money pump".

Continuity implies that no outcome is so bad that you should not be willing to take some gamble that might result in you ending up with that outcome, but might otherwise result in you ending up with a marginal improvement on your status quo, provided the chances of the better outcome are good enough. Intuitively, Continuity guarantees that an agent's evaluations of lotteries are appropriately sensitive to the probabilities of the lotteries' outcomes.

Some people find the Continuity axiom too strong. Is there any probability p such that you would be willing to accept a gamble that has that probability of you losing your life and probability $(1 - p)$ of you winning £10? (Luce & Raiffa 1989/1957, 27) Many people think there is not. However, the very same people would presumably cross the street to pick up a £10 bill they had dropped. But that is just taking a gamble that has a very small probability of being killed by a car but a much higher probability of gaining £10.

Reduction of compound lotteries is an often forgotten axiom of expected utility theory. Perhaps the reason for this is that it seems on the face of it so compelling. Informally, the axiom simply ensures that two lotteries that confer the exact same probabilities on the possible outcomes get assigned the same value. For instance, a lottery that delivers £5,000 if a fair coin comes up heads three times in a row (but otherwise delivers £0) gets assigned the same value as a lottery that delivers £5,000 if a yellow ball is randomly drawn from an urn containing seven white balls and one yellow ball. And that may seem very plausible. However, note that the axiom implies it does not matter whether the probability of an outcome is the result of the probability of a sequence of events (e.g. the coin coming up heads three times in a row) or a single event (e.g. a

[6] Here I am assuming Completeness. After all, if Completeness is not assumed, then one might violate Transitivity by weakly preferring L_j to L_i and weakly preferring L_k to L_j, while having no preference when it comes to on L_i vs. L_k.

yellow ball being drawn). And some have complained that this rules out assigning any (dis)value to gambling as such, an issue to which we shall return in section 3.3.

Independence implies that when two alternatives have the same probability for some particular outcome, our evaluation of the two alternatives should be independent of our opinion of that particular outcome. Intuitively, this means that preferences between lotteries should be governed only by the features of the lotteries that differ; the commonalities between the lotteries should be ignored. A preference ordering must satisfy some version of the Independence axiom for it to be possible to represent it as maximising what is called an *additively separable* function; for instance, a function according to which the value (i.e., expected utility) of an option is a (probability weighted) *sum* of the values of the option's possible outcomes.

To see this, suppose L_i and L_j are two alternatives, or lotteries, such that L_i will either result in outcome A, which has probability p, or C, which has probability q, and L_j will either result in outcome B, which has probability p, or C, which has probability q. Then $EU(L_i) \leq EU(L_j)$ just in case $pu(A) + qu(C) \leq pu(B) + qu(C)$. And the latter of course holds when, and only when, $pu(A) \leq pu(B)$. So an expected utility representation implies that when two alternatives have the same probability of some particular outcome, our evaluation of the two alternatives should be independent of what we think of that particular outcome, which is exactly what Independence requires.

Independence has however been extensively criticises. We shall look at that criticism in more detail in section 3.1. For now, we focus on the representation theorem that von Neumann & Morgenstern (2007/1944)'s axioms give rise to:[7]

Theorem (von Neumann-Morgenstern)**.** *Let* $\boldsymbol{O}$ *be a finite set of outcomes,* $\boldsymbol{L}$ *a set of corresponding lotteries that is closed under probability mixture and* $\precsim$ *a weak preference relation on* $\boldsymbol{L}$*. Then* $\precsim$ *satisfies axioms 1-4 if and only if there exists a function u, from* $\boldsymbol{O}$ *into the set of real numbers, that is unique up to positive linear transformation,*[8] *and relative to which* $\precsim$ *can be represented as maximising expected utility.*

One important implication of the above theorem is that, in principle at least, talk about a person's "utilities" can now be translated into talk about the person's preferences and thus her tendency to choose. Moreover, the result shows that the assumption

[7] Kreps (1988) and Peterson (2009) each provide accessible but different illustrations of how the theorem can be proven.

[8] That u is unique up to a positive linear transformation means that, for the purposes of the representation, u is considered equivalent to all and only those functions u' that satisfy $u' = a + ub$ for some number a and positive number b.

that a rational person acts so as to maximise expected utility can be stated as an assumption about the person's choice tendencies. In light of the behaviourist inclination that has dominated neoclassical economics (recall fn. 2), there is thus no wonder that economists embraced vNM's result.

For instance, note that vNM's theorem establishes that it is meaningful to ask about how the difference in utility between, say, two outcomes compares to the difference in utility between some other two outcomes. For instance, suppose we know that some agent prefers apples (A) to bananas (B) which she prefers to citrus fruit (C). We might be interested in knowing how the difference, according to her—i.e., the difference in utility—between A and B compares to the difference between B and C. And vNM's result seems to ensure that we can indeed meaningfully ask such questions.

The way to answer the above question, according to vNM's theory, is to construct a lottery between A and C, and find out what probability the lottery has to confer on A for the agent to be indifferent between on the one hand this lottery and on the other hand getting B for sure. The higher this probability, the greater the difference in utility between B and C compared to the difference between A and B. Intuitively, the higher this indifference probability, the less risks the person is willing to take in her pursuit for A rather than B when the risk can also result in C; which in turns suggests that she does not deem A a *much* better than B compared how much worse she considers C than B. For instance, if the person requires this probability to be 0.75, then that implies, by vNM's theory, that B is three quarters on the way up a utility interval that has A on the top and C on the bottom.

So, it would seem that vNM's result ensures that we can ask how the strength of a person's preference between one pair of *risk free* outcomes compares to the strength of her preference between another pair of *risk free* outcomes; and the way to answer this questions is to look at the person's preferences between *risky* lotteries. However, this inference from attitudes to risky lotteries to attitudes to risk-free outcomes has been a topic of hot debate, which will be reviewed in section 3.

2.2 Risk aversion

A noticeable feature of the expected utility equation, that has given rise to much discussion and debate, is that it assumes *risk neutrality with respect to utility*. If L_i is a *non-trivial* lottery whose expected utility is x, and L_j is a *trivial* lottery whose "expected" utility is also x, then an agent whose preferences maximise expected utility—that is, an agent whose preferences satisfy axioms 1-5—is indifferent between L_i and L_j. In other words, such a person is indifferent between any lottery whose expected utility is x and a *sure outcome* whose utility is x.

In contrast, the expected utility formula does not assume risk neutrality with respect to the outcomes to which utilities are attached. Suppose for instance that **O** is a set of possible wealth levels. Then the expected utility equation is consistent with the agent of interest being either risk averse or risk seeking with respect to wealth levels; and, in fact, consistent with the agent being risk averse when it comes to levels of wealths within some ranges while being risk seeking when it comes to levels of wealth within other ranges. Which is fortunate, since universal risk neutrality would be neither empirically nor normatively plausible.

Let's take an example to illustrate the claims in the last paragraph. Suppose that some person is offered a 50-50 gamble between winning £5,000 and losing £5,000. This lottery, or gamble, is what is called "actuarially fair": its expected monetary payoff is 0. Now let's say that the person in question has a pre-gamble wealth of w. If the person is risk *neutral* with respect to monetary amounts in the range from $w - £5,000$ to $w + £5,000$ then she is indifferent between accepting and rejecting this 50-50 gamble. However, if the person is risk *averse* when it comes to monetary amounts in this range, then she will turn down the gamble (and would continue to do so even if the potential gain were slightly increased). In contrast, if she is risk *seeking*, then she will accept the gamble (and would continue to do so even if the potential gain were slightly decreased).

Let us however focus on risk aversion. To make sense of a person turning down the 50-50 gamble between winning £5,000 and losing £5,000, within the vNM framework, we assume that the person has a utility function over quantities of money that is *concave* over the relevant interval, which means that its graph has the shape depicted in figure 1. Informally, this means that, within this range, an additional pound results in a smaller increase in utility as we move up within this range. Even less formally, this means that a pound is worth less (in utility) the more pounds the person already has; or, as it is often put, pounds have *diminishing marginal utility* (within this range). And that surely seems like a common psychological phenomena. Whether it explains risk aversion is however an issue to which we shall return in the next section.

So, the von Neumann & Morgenstern (2007/1944) framework seems to be able to account for risk aversion. And it can account for risk seeking behaviour too; for a risk seeking person, the utility function is convex rather than concave, as in the graph in

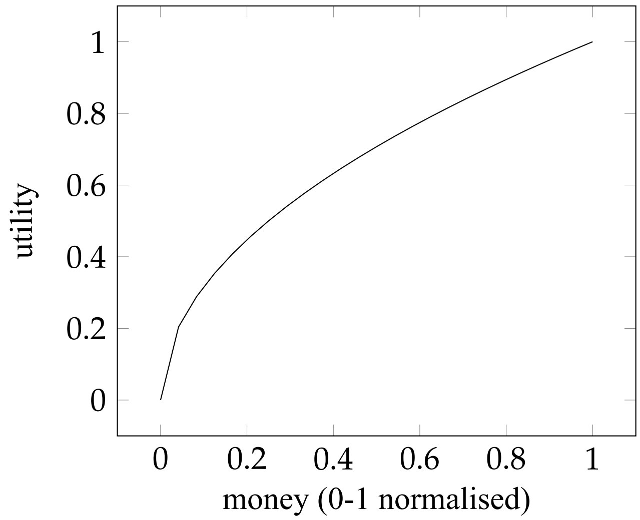

Figure 1: diminishing marginal utility

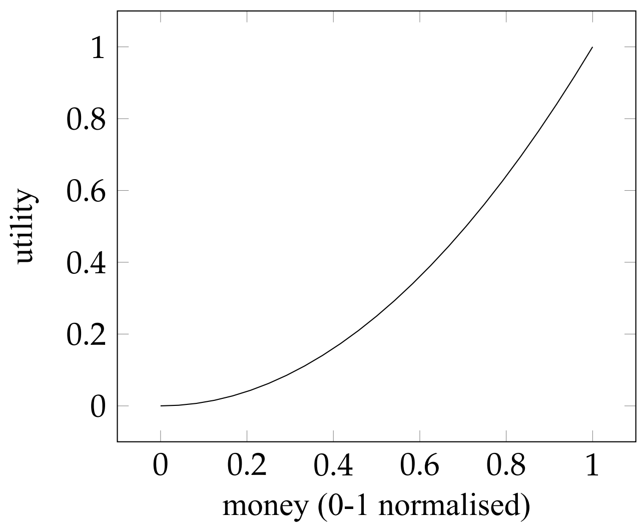

Figure 2: Increasing marginal utility

figure 2. Informally, this means that, for amounts within the relevant range, a pound is worth more (in utility) the more pounds the person already has.

Finally, the framework can account for agents who display risk seeking behaviour when it comes to monetary amounts within some ranges while displaying risk averse behaviour when it comes to amounts within other ranges. Consider for instance the fact that many people gamble in the casino—which seems to suggest risk seeking behaviour—while at the same time insuring their house—which seems to suggest risk averse behaviour. In a seminal application of expected utility theory, Friedman & Sav-

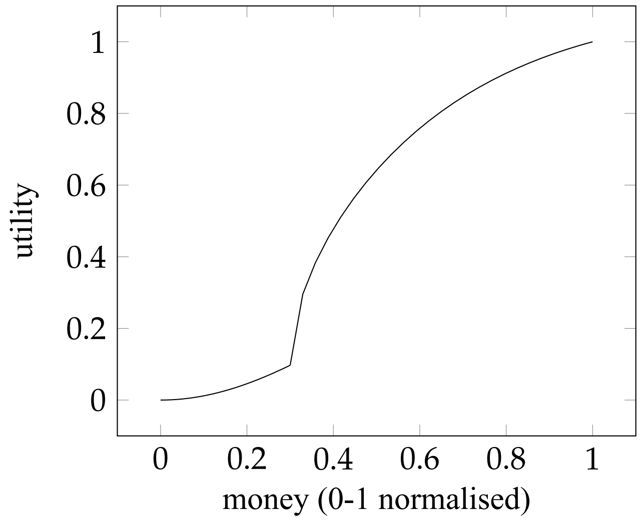

Figure 3: Decreasing and increasing marginal utility

age (1948) attempted to account for such behaviour by a utility function that has a convex shape when relative small sums of money are involved (representing risk seeking behaviour in the casino) while having a concave shape when larger sums of money are involved (representing risk averse behaviour in insurance markets). A utility function with such a shape is depicted in the graph in figure 3.

So, the framework that orthodox (neoclassical) economists use to explain, predict, and recommend choices under risk may seem flexible enough to, formally at least, represent people's differing attitudes to risk. However, in the next section we encounter some arguments purporting to show that the framework is not as flexible as its proponents have claimed.

3 Critiques of the Orthodox Treatment of Risk

Expected utility theory, as for instance developed by von Neumann & Morgenstern (2007/1944), has come under heavy criticism over the last decades. Some of this criticism is empirical, in that it uses experiments and other data to argue that people often do not act as the theory predicts. Other criticism is normative, where the complaint is that even perfectly rational people need not always act as the theory prescribes. Finally, some of the criticism is more conceptual, in that the complaint concerns attitudes or concepts that seem important for decision-making but which the theory completely ignores.

3.1 Allais' challenge

Independence is perhaps the vNM axiom that has been most critically discussed. Although the axiom seems compelling—in particular from a normative point of view—when considered in the abstract, there are famous examples where many people find that they even on reflection violate the axiom. A particularly well-known such example is the so-called *Allais Paradox*, which the French economist Allais (1953) first introduced. The paradox turns on comparing people's preferences over two pairs of lotteries similar to those given in table 1. The lotteries are described in terms of the prizes (outcomes) that are associated with particular numbered tickets, where one ticket will be drawn randomly (for instance, L_1 results in a prize of £2,500 if one of the tickets numbered 2-34 is drawn).

	1	2 – 34
L_1	£0	£2,500
L_2	£2,400	£2,400

	1	2 – 34	35 – 100
L_3	£0	£2,500	£0
L_4	£2,400	£2,400	£0

Table 1: 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 which I shall refer to as *Allais' preferences*.[9] Moreover, some scholars argue that this is a rationally permissible combination of preference. A common way to rationalise the preferences is that in the first choice situation, the risk of ending up with nothing, after choosing L_1, when one could have had £2,400 for sure, by choosing L_2, outweighs the chance that L_1 offers of a better prize (£2,500). In the second choice situation, however, the minimum one stands to gain is £0 no matter which choice one makes. Therefore, one might reason that the slight extra risk of £0 that L_3 carries over L_4 is worth taking due to L_3's chance of the better prize.

While the above reasoning may seem compelling, Allais' preferences conflict with the Independence axiom.[10] For note that in the second choice scenario option L_3 is a

[9] Kahneman & Tversky (1979) contains an influential empirical study of Allais' preferences.

[10] Thus the "paradox": many people think that Independence is a requirement of rationality, but nevertheless also think that Allais' preferences are rationally permissible.

lottery that with probability 0.34 results in lottery L_1 but that otherwise results in £0 whereas option L_4 is a lottery that with probability 0.34 results in lottery L_2 but that otherwise results in £0. So, by Independence, if one prefers L_2 to L_1 then one should prefer L_4 to L_3. And since Allais' preferences violate Independence, it follows from vNM's theorem that it cannot be represented as maximising expected utility.

There is no doubt that many people do in fact have preferences such as Allais'. Hence, some so-called *behavioural* economists have constructed decision theories that are meant to capture this type of preference without being normative, that is, without being (necessarily) intended as either guides or criteria for rational decisions. Examples of such theories include *prospect theory* (Kahneman & Tversky 1979, Tversky & Kahneman 1992), *regret theory* (Loomes & Sugden 1982, Bell 1982), and *rank-dependent utility theory* (Quiggin 1982). Since my focus here is on the orthodox (i.e., neoclassical) economic account of risk, I shall not discuss these behavioural theories in detail. An overview of descriptive decision theories can be found in Chandler (2017), while Angner (2012) is a more general introduction to behavioural economics.

Responses vary greatly when it comes to what *normative* lesson to draw from the Allais' Paradox. Leonard Savage, one of the founder of expected utility theory for subjective probabilities (Savage 1972/1954), famously failed the Allais test—that is, "failed" by the light of his own theory—but reported that having realised his mistake, he reasoned himself into agreement with the theory and thus away from the Allais' preferences (Savage 1972/1954, 101-103; for a discussion of Savage's reasoning, see Dietrich et al. 2020).

Others have argued that if it is in fact rationally permissible to take into account, when evaluating L_1, the regret or disappointment that one predicts one will experience if one gets £0 when one could have chosen £2,400 for sure, then that should somehow be accounted for in the description of L_1 (see, e.g., Weirich 1986, Broome 1991*b*). In particular, one should, according to this view, re-describe the £0 outcome of L_1 as something like "£0 + disappointment". But that makes the preference in question consistent with the Independence axiom, since L_4 is then no longer a lottery between £0 and L_1. Hence, the *paradox* might seem to have been resolved. Table 2 provides an illustration, where 'δ' stands for whatever negative feeling that one predicts one will experience if one ends up with £0 when one could have chosen £2,400 for sure.

Finally, some have argued that Allais' preferences are indeed rationally permissible *and* are better captured by some *normative* alternative to (vNM's) expected utility

	1	2 – 34
L_1	£0 + δ	£2500
L_2	£2400	£2400

	1	2 – 34	35 – 100
L_3	£0	£2500	£0
L_4	£2400	£2400	£0

Table 2: Allais' "paradox" redescribed

theory. This view is particularly popular amongst philosophers, and has for instance recently been defended by Buchak (2013) and Stefánsson & Bradley (2019). Amongst economists, the dominant view still seems to be that although we may have to depart from expected utility theory for descriptive purposes, that is, when explaining or predicting choices, expected utility theory is still unchallenged as a normative theory.

3.2 Rabin's challenge

Another well-known criticism of the descriptive accuracy of expected utility theory is based on Rabin's (2000*b*) so-called *calibration results*, the fundamental insight behind which is that expected utility theory cannot plausibly explain many people's aversion to risk when small sums of money are at stake. In short, the problem is that once a utility function has been calibrated to capture risk aversion with respect to small stakes, it will be so concave as to imply what Rabin thinks is "absurdly severe" risk aversion when more is at stake (Rabin 2000*a*).

A similar point had been made many decades earlier, by Samuelson (1963), who pointed out that that an expected utility maximiser who turns down a 50-50 gamble between wining \$200 and losing \$100 must also (to maintain consistency) turn down a bundle consisting of 100 such independent gambles. But the bundle would seem quite hard to turn down: it has a monetary expectation of \$5,000 and only has a 1/2,300 chance of resulting in the bettor losing money. "A good lawyer could have you declared legally insane for turning down this gamble," Rabin (2000*a*: 206) remarks.

Rabin in effect extended Samuelson's observation to a general calibration theorem, into which different small-scale gambles can be plugged, to see which large-scale gambles an expected utility maximiser must reject, if she rejects the inputted small-scale gambles. And the implications indeed do seem absurd. The theorem for instance establishes that an expected utility maximiser who always (i.e., irrespective of her pre-gamble wealth) turns down a 50-50 gamble between winning \$105 and losing \$100 will (if consistent) turn down a single 50-50 gamble between losing \$2,000 and winning any amount whatsoever—including an infinite amount!

Now, the above result assumes that the decision-maker turns down some particular gamble irrespective of her wealth. But Rabin's result in fact has implications even for risk averse expected utility maximisers about whom we only know that they would turn down a particular gamble when their wealth is in some particular range. For instance, the theorem implies that a risk averse expected utility maximiser who, when her pre-gamble wealth is up to \$300,000, turns down a 50-50 gamble between losing \$100 and winning \$125, would, when her wealth is no more than \$290,000, turn down a 50-50 gamble between losing \$20,000 and winning \$540,000,000,000,000,000,000!

Although Rain's results may be surprising, the logic behind the result is relatively straightforward. Recall that within expected utility theory, the form of the utility function is the only thing that can be varied to account for different attitudes to risk. In particular, risk aversion is equated with a concave utility function, or "diminishing marginal utility". And, as (Rabin 2000*b*, 1285) nicely illustrates:

> if you reject a 50-50 lose \$10/gain \$11 gamble because of diminishing marginal utility, it must be that you value the eleventh dollar above your current wealth by at most (10/11) as much as you valued the tenth-to-last-dollar of your current wealth. Iterating this observation, if you have the same aversion to the lose \$10/gain \$11 bet if you were \$21 wealthier, you value the thirty-second dollar above your current wealth by at most $(10/11) \times (10/11) \approx (5/6)$ as much as your tenth-to-last dollar. You will value your two-hundred-twentieth dollar by at most (3/20) as much as your last dollar, and your eight-hundred-eightieth dollar by at most (1/2,000) of your last dollar. This is an absurd rate for the value of money to deteriorate—and the theorem shows the rate of deterioration implied by expected-utility theory is actually quicker than this.

A natural response to Rabin's results—and, in fact, the response Rabin himself suggested (Rabin 2000*b*, 1288-1289)—is that at least when it comes to explaining people's aversion to risk when little is at stake, expected utility theory should be replaced by some theory that incorporates what is called *loss aversion*. The most important features of such theories are, first, that they incorporate some *status quo*, and define utility in terms of changes in wealth relative to this status quo rather than in terms of absolute wealth. Moreover, such theories postulate that people are more concerned by losses than with gains relative to this status quo; informally, the disutility of losing \$100 is greater than the utility of gaining \$100, relative to any status quo. Such loss aversion is one of the key ingredients of prospect theory (Kahneman & Tversky 1979), which, as previously mentioned, will not be discussed in any detail in this chapter.

3.3 Phenomenological challenges

Another common complaint against the vNM approach is that it mischaracterises attitudes to risk. Such attitudes, the complaint goes, need to be more clearly distinguished from attitudes to risk-free outcomes than the vNM approach allows. Recall that this approach *equates* different attitudes to risk with different forms of the utility function over quantities of risk free outcomes; for instance, risk aversion with respect to money is equated with diminishing marginal utility of money. A problem with this equation, according to the critics, is that attitudes to risk *per se* simply seem to be a different type of psychological attitude than attitudes to quantities of risk-fee outcomes. But, as vNM themselves pointed out, "concepts like 'specific utility of gambling' [i.e., what I called attitudes to risk *per se*] cannot be formulated free of contradiction" within their framework (von Neumann & Morgenstern 2007/1944: 28).

Critics of expected utility theory argue that, contrary to what the aforementioned equation implies, it is conceptually possible that two individuals 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, for instance due to different gambling temperaments (Watkins 1977, Hansson 1988, Buchak 2013, Stefánsson & Bradley 2019). For instance, imagine that two people both insist that they evaluate money linearly, which for instance means that the difference (in utility) between winning £50 and winning £0 is exactly as great as the difference between winning £100 and winning £50. Nevertheless, one of them is eager to accept, while the other turns down, a 50-50 gamble between winning £100 and losing £100. And the explanation they give is simply that they have different attitudes to taking risks; one of them enjoys gambling while the other detests it.

A standard response that economists have, historically at least, made when confronted with criticism like that above, is to suggest a *formalistic* interpretation of expected utility, according to which the role of expected utility theory is not to capture what actually goes on in people's minds, when making a decision, but simply to mathematically *represent* and *predict* choices (see, e.g., Friedman & Savage 1948 and Harsanyi 1977). If that is the aim, then as long as we can represent, say, a risk averse decision-maker *as if* she were maximising the expectation of some concave utility function, then it does not matter that we are conflating two conceptually distinct psychological attitudes. In other words, as long as, say, diminishing marginal utility is behaviourally indistinct from aversion to risk *per se*, it does not matter whether or not these are psychologically distinct.

The formalistic interpretation has been criticised by several philosophers of economics.[11] One complaint 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 for decision-making purposes (such as in policy analysis), we need to assume that the utilities on which we base the recommendations exist prior to (and independently of) the choices that the theory recommends. That is, if we want to be able to recommend a risky option because it is the one that maximises expected utility, then we must understand "utility" as something that is independent of the decision-makers choices—and conceptually distinct from the representation of her preferences—between risky options.

However, a proponent of the vNM theory might respond that the theory is only meant to apply to persons and situations where attitudes to risk *per se* have no influence on the person's preference. And even with that limitation, the theory is very powerful; for instance, it allows us to derive a precise utility function over quantities of goods from the persons preference between lotteries. In fact, Binmore (2009) points out that it is only *because* a vNM utility function cannot account for attitudes to risk *per se* that such a function can plausibly explain the agent's choice in situations where risk is lacking:

> It is often taken for granted that gambling can be explained [within an expected utility framework] as rational behavior on the part of a risk-loving agent. ... The mistake is easily made, because to speak of "attitude to risk" is a positive invitation to regard the shape of [a person's vNM function] as

[11] See, for instance, Broome (1991*a*), List & Dietrich (2016), Reiss (2013), Bradley (2017), and Okasha (2016).

> embodying the thrill that she derives from the act of gambling. But if we fall into this error, we have no answer to the critics who ask why [vNM functions] should be thought to have any relevance to how [the person] chooses in riskless situations. (54)

Moreover, Binmore identifies Reduction of Compound Lotteries as the reason why the vNM framework is not equipped to represent agents who are not neutral to risk *per se*.

> [Reduction of Compound Lotteries] takes for granted that [a person] is entirely *neutral* about the actual act of gambling. She doesn't bet because she enjoys betting—she bets only when she judges that the odds are in her favor. If she liked or disliked the act of gambling itself, we would have no reason to assume that she is indifferent between a compound lottery and a simple lottery in which the prizes are available with the same probabilities. (ibid, emphasis in original)

In other words, proponents of the vNM framework face a dilemma. They can accept that their framework cannot account for any potential thrill or anxiety that an agent derives from the act of gambling, that is, the framework cannot account for attitudes to risk *per se*. Or they can accept that the utility functions that the framework allows the modeller to derive cannot be used to explain or predict how the modelled agent chooses in a riskless situation. So, either the framework cannot account for attitudes to risk *per se*, or it is of little relevance to choice without risk.

4 Uncertainty

So far the focus has been on decision-making under risk, that is, situations where the agent knows—or, at least, deems suitable to act as if she knew—the relevant probabilities. Betting on a roulette is a paradigm example. In contrast, when betting on a soccer match, one does not, as previously mentioned, know all the relevant probabilities; nor would one typically find it suitable or reasonable to act as if one knew these probabilities. In the latter case, economic theorists say that the bettor is making a decision *under uncertainty*.

Leonard Savage's (1972/1954) decision theory is without a doubt the best-known normative theory of choice under uncertainty, in particular within neoclassical economics. Savage formulated a set of preference axioms that guarantee the existence of a pair of probability and utility functions relative to which the preferences can be represented as maximising expected utility. The theory is often called *subjective* expected utility theory, as the probability function is assumed to be subjective (in contrast to the previously discussed expected utility theory with "objective" or known probabilities). Since the focus of this chapter is decision-making under risk, I shall only present Savage's theory very briefly here; a somewhat more detailed account can be found in Steele and Stefánsson (2015).

The primitives in Savage's theory are outcomes (or "consequences", as Savage called them) and states of the world, the former being whatever is of ultimate value to the agent, while the latter are features of the world that the agent cannot control and about which she is typically uncertain. Sets of states are called *events*. The options over which the agent has preferences in Savage's theory are a rich set of *acts*, that formally are functions from the set of outcomes to the set of states of the world. So, an agent can choose between acts, and the outcome of an act is determined by what is the true (or the actual) state of the world.

The following notation will be used to state Savage's representation result: f, g, etc, are various acts, i.e., functions from the set **S** of states of the world to the set **O** of outcomes, with **F** being the set of these functions. $f(s_i)$ denotes the outcome of f when state $s_i \in \mathbf{S}$ is actual. The subjective expected utility of f, according to Savage's theory, denoted $U(f)$, is then given by:

Savage expected utility equation. $U(f) = \sum_i u(f(s_i)).P(s_i)$

The result Savage proved can be stated as follows:[12]

Theorem (Savage)**.** *Let $\preceq$ be a weak preference relation on **F**. If $\preceq$ satisfies Savage's axioms, then the following holds:*

- *The agent's uncertainty with respect to the states in **S** can be represented by a* unique *(and finitely additive) probability function, P;*
- *the strength of her desires for the sure outcomes in **O** can be represented by a utility function, u, that is unique up to positive linear transformation;*

[12] I assume that the set **O** is finite, but Savage proved a similar result for an infinite **O**.

- *and the pair (P, u) gives rise to an expected utility function, U, that represents her preferences for the alternatives in $\boldsymbol{F}$; i.e. for any $f, g \in \boldsymbol{F}$:*

$$f \preceq g \Leftrightarrow U(f) \leq U(g)$$

I will not present all of Savage's axioms. Instead, I focus on what is arguably the cornerstone of Savage's subjective expected utility theory, and which corresponds to von Neumann and Morgenstern's Independence axiom.

To state the axiom in question, we say that act f "agrees with" act g in event E if, for any state in event E, f and g yield the same outcome.

Axiom 6 (Sure Thing Principle)**.** *If f, g, and f', g' are such that:*

- *f agrees with g and f' agrees with g' in event $\neg E$,*
- *f agrees with f' and g agrees with g' in event E,*
- *and $f \preceq g$,*

then $f' \preceq g'$.

The idea behind the Sure Thing Principle (STP) is essentially the same as that behind Independence: since we should be able to evaluate each outcome independently of other possible outcomes, we can safely ignore states of the world where two acts that we are comparing result in the same outcome. And, for that reason, the Allais paradox—or at least some variant of it without known probabilities—is often seen as a challenge to the STP. Putting the principle in tabular form may make this more apparent. The setup involves four acts with the following form:

	E	$\neg E$
f	X	Z
g	Y	Z
f'	X	W
g'	Y	W

The intuition behind the STP is that if g is weakly preferred to f, then that must be because the consequence Y is considered at least as desirable as X, which by the same reasoning implies that g' is weakly preferred to f'.

One of the most discussed challenge to Savage's theory—in fact, a challenge to both to STP and to Savage's definition of comparative belief—is based on a choice situation devised by Daniel Ellsberg (1961). The choice situation gives rise to what is often called the *Ellsberg paradox*, since when confronted with the choices he presented, most people exhibit a pair of preferences—"Ellsberg's preferences"—that seem intuitively rational, but nevertheless conflict with Savage's theory.

Imagine an urn with 90 balls, 30 of which are red, but the remaining 60 a mix of black and yellow balls in a proportion that is unknown to the decision-maker. A ball will be randomly drawn from the urn, but first the decision-maker is offered two choices, each between a pair of bets. The four bets are presented in table3. First, she is offered a choice between bet f, which results in a prize of \$100 if a red ball is drawn (but nothing otherwise), and bet g, which pays out \$100 if a black ball is drawn (but nothing otherwise). Many people, it turns out, choose f over g. Next, the decision-maker is offered a choice between f', which results in a prize of \$100 in the event that a red or yellow ball is drawn (but nothing otherwise), and g', which pays out \$100 if a black or yellow ball is drawn (but nothing otherwise). This time many people prefer g' over f'. In fact, many people prefer both f over g and g' over f', in accordance with Ellsberg's preferences.

	red	black	yellow
f	\$100	\$0	\$0
g	\$0	\$100	\$0
f'	\$100	\$0	\$100
g'	\$0	\$100	\$100

Table 3: Ellsberg's bets

The intuitive justification for this pair of preference is that when offered the choice between f and g, people prefer the former, since they *know* that it has a 1/3 chance of resulting in them receiving \$100, whereas the chance that the second bet results in them winning \$100 can be anywhere from 0 to 2/3. The same type of reasoning would lead to a choice of g' over f': bet f' is known to have a 2/3 chance of delivering the \$100, whereas the former offers a chance anywhere between 1/3 and 1/1.

However, it is not too difficult to see that there is no single probability function over the relevant events relative to which Ellsberg's preference can be represented as maximising expected utility (assuming that table 3 correctly represent the decision problem). The problem is that if a person prefers \$100 to \$0, then, by Savage's utility representation, the first preference, $g \prec f$, reveals that the person takes it to be more probable that a red ball is drawn than that a black ball is drawn, but the second preference, $f' \prec g'$, reveals that the person takes it to be more probable that a black or yellow ball is drawn than that a red or yellow ball is drawn. But there is no probability function such that a red ball is more probable than a black ball, yet a black or yellow ball is more probable than a red or yellow ball. Hence, there is no probability function relative to which a person with Ellsberg's preferences can be represented as maximising expected utility, as defined by Savage.

It is also easy to verify that Ellsberg's preference is inconsistent with Savage's Sure Thing Principle. The principle entails that since f and g yield the same outcome in the event that a yellow ball is drawn, we can ignore this event when choosing between f and g. The same holds when choosing between f' and g'. But when we ignore this event in both choices, f becomes identical to f' and g to g'. Therefore, a preference for f over g is, according to the STP, only consistent with a preference for f' over g'. So Ellsberg's preference (f over g and g' over f') is inconsistent with the STP.

The perhaps most common rationalisation of Ellsberg's preference, at least within economics, is to suggest that people are using a *maximin expectation* rule,[13] which tells you to choose an alternative whose worst possible expectation is better than (or at least as good as) the worst possible expectation of any other alternative (Gilboa and Schmeidler 1989 axiomatised this rule and, to some extent, popularised it within economics).

Recall that in the first of Ellsberg's choice situations, the monetary expectation of betting on read is known to be \$33.33 (since one knows that 30 balls out of 90 are red). However, the monetary expectation of betting on black could be anywhere between \$0 and \$66.67. So it might make sense for a person who is *averse to uncertainty* (or averse to *ambiguity*, as it is often called) to bet on red, since it concerns no uncertainty about the expectation, which is precisely what maximin expectation prescribes. Analogous reasoning would lead to a a bet on black or yellow (bet g') in the second of Ellsberg's choice situation. So, the maximin expectation rule prescribes choices in accordance with Ellsberg's preference.

An alternative rationalisation of Ellsberg's preferences, which was recently pro-

[13]This is a variant of the maximin decision rule, which tells you to choose an alternative whose worst possible outcome is better than (or at least as good as) the worst possible outcome from any other alternative.

posed by the philosopher Richard Bradley (2016) but has not been as influential in economics, is that people with Ellsberg's preferences take quantities of chances to have decreasing marginal utility, such that, for instance, the difference in utility between no chance of \$100 and 1/3 chance of \$100 is greater than the difference in utility between 2/3 chance of \$100 and the certainty of \$100. The question of how (and, in fact, whether) to rationalise Ellsberg's preference—and, more generally, how to think of rational decisions under uncertainty—is a matter of active debate, that will not be settled here.

5 Concluding remarks

Neoclassical economists use expected utility theory to explain, predict, and guide choices in situations of risk, and the similar theory of subjective expected utility theory to explain, predict, and guide choices in situations of uncertainty. The main aim of this chapter has been to, first, describe these theories, and, second, discuss some of challenges that these theories face. Since a considerable part of the chapter has been devoted to the challenges, I would like to end on two remarks in expected utility theory's favour; remarks that both support the objective (vNM) and the subjective (Savage) versions of the theory.

First, when it comes to descriptive purposes, some economists have forcefully argued that we do not yet have a good reason for giving up on expected utility theory. The reason is that although we have found that in some experimental settings, different descriptive theories have better predictive success than expected utility theory, there is no single descriptive theory that does better than expected utility theory across these different experimental settings. Hence, some economists suggest that we should favour simplicity over complexity and stick with expected utility theory (e.g. Binmore 2009: 58-59).

Second, a forceful argument for the normative plausibility of expected utility theory comes from considerations that are similar to the money pump argument that we already encountered when discussing the axiom of Transitivity. For instance, a money-pump-like "dynamic consistency" argument can be made in favour of both Independence and the Sure Thing Principle—the axioms of respectively objective and subjective expected utility theory that has received most criticism. In particular, it can be shown that a decision-maker who violates either Independence or the Sure Thing Principle would do better, by her own lights, if she satisfied the axiom (for a recent overview, see Gustafsson ta.). In fact, expected utility theory as a whole can be derived from what may seem to be nothing but dynamic consistency constraints (Hammond 1987, 1988). So, while some think that examples such as the paradoxes of Allais and Ellsberg undermine the normative standing of expected utility theory, we still have compelling dynamic and practical arguments in favour of the theory.[14]

BIOGRAPHICAL NOTE H.Orri Stefánsson is associate professor of practical philosophy at Stockholm University, Pro Futura Scientia Fellow at the Swedish Collegium for Advanced Study, and advisor at the Institute for Futures Studies. He works mostly on decision theory and related topics, with a current focus on decision-making under extreme uncertainty.

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.

Angner, E. (2012), *A Course in Behavioral Economics*, Palgrave Macmillan.

Angner, E. & Loewenstein, G. (2012), Behavioral economics, *in* U. Mäki, ed., 'Philosophy of Economics', North-Holland, pp. 641–689.

Aumann, R. J. (1962), 'Utility theory without the completeness axiom', *Econometrica* **30**(3), 445–462.

Bell, D. E. (1982), 'Regret in decision making under uncertainty', *Operations Research* **30**(5), 961–981.

Binmore, K. (2009), *Rational Decisions*, Princeton University Press.

Bradley, R. (2001), Ramsey and the measurement of belief, *in* D. Corfield & J. Williamson, eds, 'Foundations of Bayesianism', Springer.

Bradley, R. (2016), 'Ellsberg's paradox and the value of chances', *Economics and Philosophy* **32**(2), 231–248.

Bradley, R. (2017), *Decision Theory With a Human Face*, Oxford University Press.

Broome, J. (1991*a*), 'Utility', *Economics and Philosophy* **7**(1), 1–12.

Broome, J. (1991*b*), *Weighing Goods*, Basil Blackwell.

[14]Financial support from Riksbankens Jubileumsfond (through a Pro Futura Scientia XIII fellowship) is gratefully acknowledged.

Buchak, L. (2013), *Risk and Rationality*, Oxford University Press.

Chandler, J. (2017), Descriptive Decision Theory, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', winter 2017 edn, Metaphysics Research Lab, Stanford University.

Davidson, D., McKinsey, J. C. C. & Suppes, P. (1955), 'Outlines of a formal theory of value, I', *Philosophy of Science* **22**(2), 140–160.

Dietrich, F., Staras, A. & Sugden, R. (2020), 'Savage's response to allais as Broomean reasoning', *Journal of Economic Methodology* **0**(0), 1–22.

Ellsberg, D. (1961), 'Risk, ambiguity, and the Savage axioms', *Quarterly Journal of Economics* **75**(4), 643–669.

Friedman, M. (1953), The methodology of positive economics, *in* 'Essays in Positive Economics', University of Chicago Press, pp. 3–43.

Friedman, M. & Savage, L. J. (1948), 'The utility analysis of choices involving risk', *Journal of Political Economy* **56**(4), 279–304.
URL: *http://www.jstor.org/stable/1826045*

Frigg, R. & Hoefer, C. (2010), Determinism and chance from a Humean perspective, *in* Dennis Dieks et.al, ed., 'The Present Situation in the Philosophy of Science', Springer, pp. 351–271.

Gilboa, I. (2009), *Theory of Decision under Uncertainty*, Cambridge University Press.

Gilboa, I. & Schmeidler, D. (1989), 'Maxmin expected utility with non-unique prior', *Journal of Mathematical Economics* **18**(2), 141–153.

Gustafsson, J. E. (ta.), *Money-Pump Arguments*, Cambridge University Press.

Hammond, P. (1987), Consequentialism and and the Independence axiom, *in* B. Munier, ed., 'Risk, Decision and Rationality', Springer.

Hammond, P. (1988), 'Consequentialist foundations for expected utility theory', *Theory and Decision* **25**(1), 25–78.

Hansson, B. (1988), Risk aversion as a problem of conjoint measurement, *in* P. Gärdenfors & 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 & J. Hintikka, eds, 'Foundational Problems in the Special Sciences', D. Reidel Publishing.

Hoefer, C. (2007), ‘The third way on objective proabability: A sceptic’s guide to objective chance’, *Mind* **116**(463), 449–496.

Jeffrey, R. (1990/1965), *The Logic of Decision*, The University of Chicago Press.

Kahneman, D. & Tversky, A. (1979), ‘Prospect theory: An analysis of decision under risk’, *Econometrica* **47**(2), 263–292.

Knight, F. H. (1921), *Risk, Uncertainty and Profit*, Houghton Mifflin Company.

Kreps, D. (1988), *Notes on the Theory of Choice*, Avalon Publishing.

List, C. & Dietrich, F. (2016), ‘Mentalism versus behaviourism in economics: A philosophy-of-science perspective’, *Economics and Philosophy* **32**(2), 249–281.

Loomes, G. & Sugden, R. (1982), ‘Regret theory: An alternative theory of rational choice under risk’, *The Economic Journal* **92**, 805–824.

Luce, R. D. & Raiffa, H. (1989/1957), *Games and Decisions: Introduction and Critical Survey*, Dover Publications.

Okasha, S. (2016), ‘On the interpretation of decision theory’, *Economics and Philosophy* **32**(3), 409–433.

Peterson, M. (2009), *An Introduction to Decision Theory*, Cambridge University Press.

Quiggin, J. (1982), ‘A theory of anticipated utility’, *Journal of Economic Behavior & Organization* **3**(4), 323–343.

Rabin, M. (2000*a*), Diminishing marginal utility of wealth cannot explain risk aversion, *in* D. Kahneman & A. Tversky, eds, ‘Choices, Values and Frames’, Cambridge University Press.

Rabin, M. (2000*b*), ‘Risk aversion and expected utility theory: A calibration theorem’, *Econometrica* **68**(5), 1281–1292.

Rachels, S. (1998), ‘Counterexamples to the transitivity of better than’, *Australasian Journal of Philosophy* **76**(1), 71–83.

Ramsey, F. P. (1990/1926), Truth and probability, *in* D. H. Mellor, ed., ‘Philosophical Papers’, Cambridge University Press.

Reiss, J. (2013), *Philosophy of Economics: A Contemporary Introduction*, Routledge.

Robbins, L. (1932), *An essay on the nature and significance of economic science*, Macmillan & Co.

Samuelson, P. A. (1938), ‘A note on the pure theory of consumer’s behavior’, *Economica* **5**, 61–71.

Samuelson, P. A. (1948), ‘Consumption theory in terms of revealed preference’, *Economica* **15**, 243–253.

Samuelson, P. A. (1963), ‘Risk and uncertainty: A fallacy of large numbers’, *Scientia* (98), 108–113.

Savage, L. (1972/1954), *The Foundations of Statistics*, Dover Publication.

Schipper, B. C. (2015), Awareness, *in* H. van Ditmarsch, J. Y. Halpern, W. van der Hoek & B. Kooi, eds, ‘Handbook of Epistemic Logic’, College Publications, pp. 77–146.

Steele, K. & Stefánsson, H. O. (2015), Decision theory, *in* E. Zalta, ed., ‘Stanford Encyclopedia of Philosophy’.

Steele, K. & Stefánsson, H. O. (2021), *Beyond Uncertainty: Reasoning with Unknown Possibilities*, Cambridge University Press.

Stefánsson, H. O. & Bradley, R. (2019), ‘What is risk aversion?’, *British Journal for the Philosophy of Science* **70**(1), 77–102.

Temkin, L. (1987), ‘Intransitivity and the mere addition paradox’, *Philosophy and Public Affairs* **16**(2), 138–187.

Temkin, L. (1996), ‘A continuum argument for intransitivity’, *Philosophy and Public Affairs* **25**(3), 175–210.

Temkin, L. (2012), *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*, Oxford University Press.

Tversky, A. & Kahneman, D. (1992), ‘Advances in prospect theory: Cumulative representation of uncertainty’, *Journal of Risk and Uncertainty* **5**(4), 297–323.

von Neumann, J. & Morgenstern, O. (2007/1944), *Theory of Games and Economic Behavior*, Princeton University Press.

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

Weirich, P. (1986), ‘Expected utility and risk’, *British Journal for the Philosophy of Science* **37**(4), 419–442.