

## **DECISION THEORY**

James M. Joyce  
Department of Philosophy  
The University of Michigan

Decision theory seeks to provide a normative account of rational decision making, and to determine the extent to which human agents succeed in living up to the rational ideal. Though many decision theories have been proposed, the version of *expected utility theory* developed in L. J. Savage's (1954/1972) classic *Foundations of Statistics* remains the best developed and most influential. It will serve as the principal focus of this entry. Savage established a general framework for thinking about decision problems. He codified core tenets of the theory of rational preference and argued cogently for them. Most important, he proved a *representation theorem* that helps to legitimize subjective Bayesian approaches to epistemology and to justify *subjective expected utility maximization* as the foundation of rational decision making (see BAYESIANISM). Indeed, Savage's contributions are so seminal that the best way to approach the topic of decision theory is by treating his theory as a kind of a "standard model," and discussing other views as reactions or additions to it. This is the approach taken here.

The entry has three sections. The first section discusses the general notion of a decision problem. The second introduces the expected utility hypothesis, and explains Savage's representation theorem. The third presents the "standard" theory of rational preference, and discusses objections to it.

### **Decision Problems**

Savage's model assumes a rational decision-maker, hereafter the *agent*, who uses beliefs about possible *states of the world* to choose *actions* that can be expected to produce desirable *consequences*. The states describe all relevant contingencies that lie beyond the agent's direct control. Any uncertainty that figures into the agent's choice is portrayed as ignorance about which state obtains. *Events* are disjunctions of states that provide less specific descriptions than states do of the circumstances under which choices are made. Consequences serve as objects of *non-instrumental* desire. Each specifies a possible course of events that is sufficiently detailed to settle every matter about which the agent intrinsically cares. Acts are objects of *instrumental* desire; the agent values them only insofar as they provide a means to the end of securing desirable consequences. When there are only finitely many acts and states the person's choice can be described using a *decision matrix*

	$S_1$	$S_2$	$S_3 \dots$	$S_n$
$A_1$	$C_{1,1}$	$C_{1,2}$	$C_{1,3 \dots}$	$C_{1,n}$
$A_2$	$C_{2,1}$	$C_{2,2}$	$C_{2,3 \dots}$	$C_{2,n}$
$A_3$	$C_{3,1}$	$C_{3,2}$	$C_{3,3 \dots}$	$C_{3,n}$
:	:	:	:	:
$A_m$	$C_{m,1}$	$C_{m,2}$	$C_{m,3 \dots}$	$C_{m,n}$

where the  $S_j$ s are states, the  $A_i$ s are acts, and  $C_{i,j}$  is the outcome that  $A_i$  will produce when  $S_j$  obtains.

This model of decision problems applies to "one-choice" decisions made at a specific time. Though early decision theorists, like Savage, believed that sequences of decisions could be

represented by one-shot choices among contingency plans, or *strategies*, this view now has few adherents. The topic of dynamic decision-making lies beyond the scope of this entry. For relevant discussions see Hammond (1988), McClennen (1990), and Levi (1991).

A decision problem is only counted as rational if the following conditions hold:

- The value of each consequence  $C$  is independent of the act and state that bring it about.
- Each act/state pair  $(A, S)$  determines a unique consequence  $C_{A,S}$ .
- The agent cannot causally influence which state obtains.

Many misguided objections to expected utility theory involve decision problems that violate these requirements. For example, the following is a central tenet of the theory

*Comparative Probability (CP)*. For any event  $E$  and consequences  $C$  and  $C^*$  with  $C$  preferred to  $C^*$ , the agent prefers an act that produces  $C$  when  $E$  and  $C^*$  when  $\neg E$  to an act that produces  $C^*$  when  $E$  and  $C$  when  $\neg E$  if and only if the agent is more confident in  $E$  than in  $\neg E$ .

This seems susceptible to counterexample. Imagine a person who is convinced that the average annual inflation rate over the next decade will be either 10% or 1%, and who thinks 10% is far more likely.

$E =$ The rate of inflation will be high over the next decade.	The rate of inflation will not be high over the next decade.
--	--

<b>A</b>	Be paid \$1000 in ten years.	Be paid \$0 in ten years.
<b>A*</b>	Be paid \$0 in ten years.	Be paid \$1000 in ten years.

Despite confidence in *E*, such a person may prefer *A\** to *A* on the grounds that \$1000 will be worth more if inflation is low than if it is high. This preference does not refute **CP**. **CP** only applies to preferences in well-formed decision problems, and this problem as *ill-formed* since the consequence “Be paid \$1000 in ten years” is worth less when it appears in the upper left than when it appears in the lower right. (Proponents of *state-dependent utility theory* relax this requirement by allowing utilities of outcomes to vary with states. See Karni (1985).) To fix the problem one needs to rewrite outcomes as follows:

	The rate of inflation will be high over the next decade.	The rate of inflation will not be high over the next decade.
<b>A</b>	Be paid \$1000 after ten years of high inflation.	Be paid \$0 after ten years of low inflation
<b>A*</b>	Be paid \$0 after ten years of high inflation.	Be paid \$1000 after ten years of low inflation.

When the problem is redescribed this way the preference for *A* over *A\** does not violate **CP**.

Similar problems arise when the decision-maker can influence states of the world.

Another core tenet of expected utility theory is

*Dominance.* If the agent prefers  $A$ 's consequences to  $A^*$ 's consequences in every possible state of the world, then the agent prefers  $A$  to  $A^*$ .

Dominance sometimes seems to make absurd recommendations.

	<b>One will contract influenza this winter.</b>	<b>One will not contract influenza this winter.</b>
<b>Get a flu shot.</b>	Get the flu, and suffer the minor pain of a shot.	Avoid the flu, but suffer the minor pain of a shot.
<b>Do not get a flu shot.</b>	Get the flu, but avoid the minor pain of a shot.	Avoid the flu, and avoid the minor pain of a shot.

Here it seems as if Dominance requires one to forgo the shot to avoid the pain, which is terrible advice given that the chances of getting the flu are markedly less with the shot than without it. Again, this is not an objection to expected utility theory, but an ill-posed decision problem. To properly reformulate the problem one must use states like these:

- ◆ One will contract the flu whether or not one gets the shot.
- ◆ One will not contract the flu whether or not one gets the shot.
- ◆ One will contract the flu if one gets the shot, but not otherwise.
- ◆ One will not contract the flu if one gets the shot, but will otherwise.

Dominance reasoning always holds good for states that are independent of the agent's acts, as these states are.

The debate between *causal* and *evidential* decision theorists has to do with the sort of independence that is required here. Evidentialists believe that states must be *evidentially* independent of acts, so that no act provides a sign or signal of the occurrence of any state. Causal decision theorists adopt the stronger requirement that states must be *causally* independent of acts, so that nothing the agent can do will change the probabilities of states. See Jeffrey (1983), Gibbard and Harper (1978), Skyrms (1980), and Joyce (1999) for details.

### **Expected Utility Representations of Preference**

Following Savage, it is standard in decision theory to assume that, after due deliberation, a rational agent will be able to order acts with respect to their effectiveness as instruments for producing desirable outcomes. This generates a *weak preference ranking*  $A \geq A^*$  that holds between acts  $A$  and  $A^*$  just when, all things considered, the agent strictly prefers  $A$  to  $A^*$  or is indifferent between them. It is important to understand that this preference ranking among acts is *not* what the agent *starts out* with when making her decision. It is the *end result* of the deliberative process.

Decision theorists have historically understood preferences behavioristically, so that  $A > A^*$  *means* that the agent would choose  $A$  over  $A^*$  if given the chance. Though some social scientists still adhere to this interpretation, it has been widely and effectively criticized. The

alternative is to take preferences as representing one's *all-things-considered judgments* about which acts will best serve one's interests. On this reading, saying one prefers  $A$  to  $A^*$  means that the balance of one's reasons favors realizing  $A$  rather than  $A^*$ , whether one can or will choose  $A$  is another matter.

An act  $A$  is *choiceworthy* when it is weakly preferred to every alternative. According to the expected utility hypothesis, rationally choiceworthy acts maximize the decision-maker's *subjective expected utility*. In Savage's framework, an act's expected utility is defined as

$$Exp_{\mathbf{P},\mathbf{U}}(A) = \sum_S \mathbf{P}(S) \times \mathbf{U}(C_{A,S})$$

where  $\mathbf{P}$  is a *probability function* defined over events, and  $\mathbf{U}$  is a real-valued *utility function* defined over consequences. To show that rationally choiceworthy acts maximize expected utility, Savage imposed a system of axiomatic rationality constraints on preference rankings, and then proved that any ranking satisfying his axioms would be consistent with the hypothesis that acts are ranked according to expected utility. (The first result of this type is found in Ramsey (1931).) One can think of Savage as seeking to establish the following two claims:

*Theory of Practical Rationality.* Any practically rational agent will have a preference ranking among acts that obeys Savage's axioms.

*Existence of Subjective Expected Utility Representations.* For any preference ranking that obeys Savage's axioms there will be at least one probability/utility pair  $(\mathbf{P}, \mathbf{U})$  such that

- $\mathbf{P}$  represents the agent's beliefs: one event  $E$  is taken to be at least as likely as another  $E^*$  only if  $\mathbf{P}(E) \geq \mathbf{P}(E^*)$ .
- $\mathbf{U}$  represents the agent's (intrinsic) desires for consequences:  $C$  is weakly preferred to  $C^*$  only if  $\mathbf{U}(C) \geq \mathbf{U}(C^*)$ .
- $Exp_{\mathbf{P},\mathbf{U}}$  accurately represents the agent's (instrumental) desires for actions:  $A$  is weakly preferred to  $A^*$  only if  $Exp_{\mathbf{P},\mathbf{U}}(A) \geq Exp_{\mathbf{P},\mathbf{U}}(A^*)$ .

It follows directly that one whose preferences satisfy Savage's axioms will always behave *as if* one were choosing acts based on their expected utility (though it is no way required that one actually use this method in making the decision).

Savage also proves that this representation is unique once a unit and zero point for measuring utility have been fixed. To establish uniqueness Savage was forced to assume that the agent has determinate preferences over an extremely rich set of options. Many decision theorists reject these "richness assumptions," and so believe that an agent's beliefs and desires should be represented by *sets* of probability/utility pairs, as in Joyce (1999, pp. 102-105).

### **The Theory of Rational Preference**

Since there is no question about the validity of Savage's representation theorem, his case for expected utility maximization rests on the plausibility of his axioms as requirements of practical rationality. Rather than trying to formulate things as Savage does, it will be better to

discuss informal versions of those of his axioms and auxiliary assumptions that are components of every expected utility theory.

### **Frame Invariance and Value Independence**

It will serve to begin by considering two principles that are left implicit in most formulations of expected utility theory.

*Frame Invariance (INV)*. A rational agent's preferences among acts should depend only on the consequences the acts produce in various states of the world, and not on the ways in which these consequences, or the acts themselves, happen to be described.

*Value Independence (VALUE)*. A rational agent endows each act with a value that is independent of the decisions in which it figures.

While **INV**'s credentials as a requirement of rationality have never been seriously questioned, a great deal of empirical research suggests that people's preferences do often depend on the way in which decisions are "framed". For example, when presented with the following two decisions (see Tversky and Kahneman (1986)):

- one will be paid \$300 to choose between (a) getting another \$100 for sure or (b) getting another \$200 with probability 1/2 and \$0 with probability 1/2.
- one will be paid \$500 to make a choice between either (a\*) paying back \$100 for sure or (b\*) paying back \$0 with probability 1/2 and paying back \$200 with probability 1/2.

a surprising number of people prefer (a) in the first choice and (b\*) in the second, thus violating **INV**. The different descriptions of the same options lead people to view their choices from different perspectives. In the first case, they see themselves as having \$300 and try to *improve* their lot by choosing between (a) and (b), while in the second they (wrongly) see themselves having \$500 and try to *preserve* their fortune by choosing between (a\*) and (b\*). This generates problems when conjoined with the following facts about human behavior (see Kahneman and Tversky (1979) and Shafir and Tversky (1995)):

*Divergence from the Status Quo.* People are more concerned with *gains* and *losses*, seen as *additions* or *subtractions* from the *status quo*, than with total well-being or overall happiness.

*Asymmetrical Risk Aversion.* People tend to be risk-averse when pursuing *gains*, but risk-seeking when avoiding losses.

People who choose between (a) and (a\*) see \$300 the “status quo,” and thus prefer the less risky (a) since they are risk-averse when pursuing gains. When choosing between (b) and (b\*) they

see \$500 as the status quo, and prefer the more risky ( $b^*$ ) because they are risk-seeking when aiming to avoid losses.

**VALUE** has a number of important implications. First, it entails that one should be able to experimentally “elicit” a person’s preference between  $A$  and  $A^*$  using any of the following methods:

- *Fair Prices.* Have an agent put a “fair price” on each action, and conclude that the higher priced act is preferred.
- *Choice.* Have an agent choose between  $A$  and  $A^*$ , and conclude that the chosen act is not dispreferred.
- *Rejection.* Have an agent reject  $A$  or  $A^*$ , and conclude that the rejected act is not preferred.
- *Exchange.* Award an agent  $A^*$ , offer to trade  $A$  for  $A^*$  plus a small fee, and conclude that  $A$  is preferred if the agent makes the trade.

Surprisingly, these procedures can all yield different results, a fact that creates havoc for behaviorist analyses of preference. In cases of *preference reversal* an agent who sets a higher price on  $A$  also selects  $A^*$  in a straight choice. Shafir (1993) gives examples in which subjects choose  $A$  over  $A^*$  and yet reject  $A$  for  $A^*$ . This happens because they focus more on comparisons among “positive” features of options when choosing, but more on negative features when rejecting. When  $A$  has both more pronounced positive features and more pronounced negative features it can be both chosen and rejected. There are even cases in which an agent will

refuse to trade  $A$  for  $A^*$  and refuse to trade  $A^*$  for  $A$ . The mere fact that the person “owns” a prospect seems to make it more valuable to her. (This is referred to as *loss aversion*.)

**VALUE** also says that an agent’s preferences among options should not depend on what other options happen to be available. This entails the following two principles Sen (1971). (Note that these do *not* apply when the addition or deletion of  $A^{**}$  provides relevant information about the desirability of  $A$  or  $A^*$ .)

- Principle- $\alpha$ : If the agent will choose  $A$  over  $A^*$  and  $A^{**}$ , then the agent will choose  $A$  over  $A^*$  even when  $A^{**}$  is not available.
- Principle- $\beta$ : If the agent will choose  $A$  in a straight choice between it and  $A^*$ , then the agent will also choose  $A$  in a choice between  $A$ ,  $A^*$  and some third option that is inferior to  $A^*$ .

Actual choosers often violate these principles. As salespeople have long known, one can more easily convince a person to buy a product by offering an inferior product at a higher price. Likewise, offering too many good options can lead a person to refrain from buying products that would have purchased had the list of options had been smaller. A disconcerting violation of both principles is the finding of Redelmeier and Shafir (1995) that physicians are *less* likely to prescribe pain medication to patients when they can choose between ibuprofen and the inferior piroxicam than when they can only choose ibuprofen. While none of these empirical results have led decision theorists to question the normative standing of **VALUE**, they clearly show that expected utility theory is not an accurate *description* of human behavior.

## Completeness

Among the principles that expected utility theorists generally state as axioms are Dominance and Comparative Probability, as well as

*Transitivity.* If the agent (strictly or weakly) prefers  $A$  to  $A^*$  and  $A^*$  to  $A^{**}$ , then the agent prefers  $A$  to  $A^{**}$ .

*Completeness (COM).* The agent either strictly prefers  $A$  to  $A^*$ , strictly prefers  $A^*$  to  $A$ , or is indifferent between them.

*The “Sure-thing” Principle (STP).* If  $A$  and  $A^*$  produce the same consequences in every state consistent with an event  $\neg E$ , then the agent’s preference between  $A$  and  $A^*$  depends exclusively only on their consequences when  $E$  obtains.

Though all five axioms are controversial in various ways, **COM** and **STP** have been the most contentious.

Completeness can fail in two ways. First, an agent might have no definite preference between two prospects either because the agent’s intrinsic desires are vague or indeterminate, or because the agent has insufficient information to judge which act will better promote desirable consequences. Being “indifferent” is not the same as “having no preference.” One who is indifferent between two options judges that they are equally desirable, but one who lacks a clear

preference is unable to judge that one option is better than the other, or even that they are equally good – the person simply has no view about their relative desirabilities. Most decision theorists now admit that there is nothing irrational about having a “gappy” preference ranking, and it is becoming standard to treat **COM** (and any other axiom that requires the *existence* of preferences) as a *requirement of coherent extendibility*, see Jeffrey (1983), Kaplan (1983), Joyce (1999, pp. 103-105). On this reading, it is only irrational to have an incomplete preference ranking if it cannot be extended to a complete ranking that obeys all the other axioms.

A more serious objection to **COM** comes from those who hold that rational agents may be unable to compare acts or consequences not because of any vagueness in their beliefs or desires, but because they regard the values of these prospects as genuinely *incommensurable*, see Raz (1986) and Anderson (1993). On one version of the view, a rational agent might regard distinctive standards of evaluation as appropriate to different sorts of prospects, and so see prospects that do not fall under a common standard as incomparable. For example, a person might see it as perfectly appropriate to set a monetary price on a share of stock, but also regard it as improper to put a price on spending an afternoon with one’s children. If this is so, then a person’s preference ranking will not compare these prospects, and no extension of it consistent with that person’s values will do so either. The incommensurability debate is too involved to pursue here. The heart of the issue has to do with the ability of rational agents to “balance off” reasons for and against an option so as to come to an “all-things-considered” judgment about its desirability. Utility theorists think that such a balancing of reasons is always possible. Proponents of incommensurability deny this.

## The Sure-Thing Principle

The Sure-thing Principle forces preferences to be *separable across events*, so that a rational agent's preference between  $A$  and  $A^*$  depends *only* on what happens in states of the world in which these prospects produce *different* outcomes. When there are three states to be considered, **STP** requires an agent facing the following decision to prefer  $A$  over  $A^*$  if and only if the agent also prefers  $B$  over  $B^*$ .

	$S_1$	$S_2$	$S_3$
$A$	$C_1$	$C_2$	$C_3$
$A^*$	$C_1^*$	$C_2^*$	$C_3$
$B$	$C_1$	$C_2$	$D_3$
$B^*$	$C_1^*$	$C_2^*$	$D_3$

Formatted

In deciding between  $A$  and  $A^*$  or between  $B$  and  $B^*$ , **STP** tells the agent to *ignore* what happens when  $S_3$  holds since the same result occurs under  $S_3$  whichever option is chosen. In effect, the requirement is that the agent should be able to form a preference between the following two *act types* whether or not the value of  $x$  is known.

	$S_1$	$S_2$	$S_3$
$X$	$C_1$	$C_2$	$x$
$X^*$	$C_1^*$	$C_2^*$	$x$

**STP** has generated more controversy than any other tenet of expected utility theory.

Much of the discussion concerns two putative counterexamples, the *Allais* and *Ellsberg Paradoxes*, which seem to show that an important component of rational preference, the amount of risk or uncertainty involved in an option, is non-separable in the sense required by **STP**. In the jargon of economists, a prospect involves *risk* when the agent knows the objective probability of each state of the world. It involves *uncertainty* when the agent does not have sufficient information to assign objective probabilities to states. **STP** entails that an agent's attitudes toward risk and uncertainty can be fully captured by the combination of the agent utility function for outcomes, and the probabilistic averaging involved in the computation of expected utilities. The Allais and Ellsberg paradoxes appear to show that the theory is wrong about this.

In *Allais' paradox* (1990) agents choose between *A* and *A\** and then between *B* and *B\** (where known probabilities of states are listed).

	<b>0.33</b>	<b>0.01</b>	<b>0.66</b>
<b>A</b>	\$2500	\$0	\$2400
<b>A*</b>	\$2400	\$2400	\$2400
<b>B</b>	\$2500	\$0	\$0
<b>B*</b>	\$2400	\$2400	\$0

Most people violate **STP** by preferring *A\** to *A* and *B* to *B\**, and these preferences remain stable upon reflection. The thinking seems to be that in the first choice one should 'play it safe' and take the sure \$2400 since a 0.33 chance at an extra \$100 does not compensate for a 0.01 risk of ending up with nothing. On the other hand, since one will probably end up with nothing in the

second choice the chance of getting an extra \$100 makes the risk worth taking. Thus, Allais choosers think (a) that there is more risk involved in choosing in  $A$  over  $A^*$  than in choosing  $B$  over  $B^*$ , and (b) that this added risk justifies their non-separable preferences.

In *Ellsberg's Paradox* (1961), a ball is drawn at random from an urn that is known to contain 30 red balls, and 60 balls that are either white or blue but in unknown proportion. The agent is asked to choose between  $A$  and  $A^*$  and then between  $B$  and  $B^*$ .

	Red	White	Blue
$A$	\$100	\$0	\$0
$A^*$	\$0	\$100	\$0
$B$	\$100	\$0	\$100
$B^*$	\$0	\$100	\$100

Most people prefer  $A$  to  $A^*$  and  $B^*$  to  $B$ . People tend to prefer risk to equivalent levels of uncertainty when they have something to gain, and to prefer uncertainty to risk when they have something to lose. Thus,  $A$  is preferred to  $A^*$  because it has \$100 riding on a prospect of known risk 0.33, while  $A^*$  has that same sum riding on an uncertainty (ranging between risk 0 and risk 0.66). Likewise,  $B^*$  is preferred to  $B$  because it offers a definite 0.66 risk of \$100 where  $B$  only offers an uncertainty (ranging between risk 0.33 and risk 1.0).

Some decision theorists take the Allais and Ellsberg paradoxes to show that expected utility theory is incapable of capturing rational attitudes toward risk. The problem with **STP**, they say, is that a rational agent need not be able to form any definite preference between the act

types  $X$  and  $X^*$  because information about  $x$ 's value might provide information about the relative *risk* of options, and this information can be relevant to her preferences.

Many expected utility theorists respond to this objection by arguing that the Allais and Ellsberg paradoxes are *underdescribed*, see Broome (1991, pp. 95-115). One can render the usual Allais preferences consistent with **STP** by rewriting outcomes as follows:

	<b>0.33</b>	<b>0.01</b>	<b>0.66</b>
<b>A</b>	\$2500	\$0 instead of a sure \$2400 with $A^*$	\$2400
<b>A*</b>	\$2400	\$2400	\$2400
<b>B</b>	\$2500	\$0 instead of a probable \$0 with $B^*$	\$0
<b>B*</b>	\$2400	\$2400	\$0

If the agent prefers the second outcome in  $A$  to the second outcome in  $B$ , then there is no violation of **STP**. Moreover, there is a plausible psychological explanation for this preference. A person who ends up with \$0 when they could have had a sure \$2400 might experience pangs of *regret* that would not be felt if that person thought that ending up with \$0 was likely anyhow. Thus, the person's decision really looks like this:

	<b>0.33</b>	<b>0.01</b>	<b>0.66</b>
<b>A</b>	\$2500	\$2400 and pangs of regret	\$0
<b>A*</b>	\$2400.	\$2400	\$2400
<b>B</b>	\$2500	\$0 with little regret	\$0
<b>B*</b>	\$2400	\$0	\$2400

The Ellsberg paradox can be handled similarly. If the agent feels a special sort of “discomfort” when gains ride on uncertain prospects (or losses ride on risky prospects), then the correct description of her problem might really be

	<b>Red</b>	<b>White</b>	<b>Blue</b>
<b>A</b>	\$100	\$0	\$0
<b>A*</b>	\$0	\$100 and discomfort	\$100
<b>B</b>	\$100 and discomfort	\$0	\$100 and discomfort
<b>B*</b>	\$0	\$100	\$100

Again, there is no violation of **STP** here.

This way of eliminating counterexamples to **STP** worries many people since it looks like an expected utility theorist can *always* use it. The fact that one can *always* explain away any seeming counterexamples to expected utility theory by redescribing outcomes and postulating the necessary beliefs and desires seems to show that the theory is contentless. This objection is especially effective against behaviorist interpretations of preference. Since behaviorists can only appeal to overt choices to isolate preferences, they have no principled way of distinguishing legitimate from *ad hoc* redescriptions of decision problems. Nothing in an Allais chooser’s behavior, for example, indicates whether the agent is seeking to avoid some (unobservable) feeling of regret or is acting on the basis of non-separable attitudes toward risk.

The objection is less effective when preferences are understood as all-things-considered judgments, for it is then possible to argue that certain redescriptions are correct because they *best*

*explain* the totality of the person's behavior. If the hypothesis that people experience regret explains a great deal of human behavior, aside from violations of **STP**, then it is legitimate to use it to explain the common Allais preferences. Consider an analogy: It could be claimed that Newtonian mechanics is empty because (as is true) any pattern of observable motions can be made consistent with Newton's laws by positing the right constellation of forces. What makes this objection unconvincing is the fact that Newton was able to account for a vast array of distinct motions using the single force of gravity. The same might be true in decision theory. If it can be shown that a small number of relatively simple psychological mechanisms, including feelings of regret or discomfort in situations of risk, explain a great deal of human behavior then the *best explanation* for the Allais and Ellsberg choices might be the ones the expected utility theorists propose. Of course, this places a burden on these theorists to show that, by standard canons of scientific reasoning, their explanation is indeed the best available.

An alternative response is to take the description of the Allais and Ellsberg paradoxes at face value, and to argue that the common preferences are irrational. To see how the argument might go for the former, note that the common rationale for the Allais choices assumes that the difference in risk between  $A$  and  $A^*$  exceeds the difference in risk between  $B$  and  $B^*$ . Proponents of expected utility theory will argue that this is mistaken. The best way to determine how much two options differ in risk, they will claim, is to ask how one might insure against the increased chances of loss that one assumes in exchanging the less risky option for the more risky one. Someone who switches from  $A^*$  to  $A$  in the Allais Paradox can insure against the risk of loss by buying an insurance policy that pays out \$2400 contingent on the 0.01 probability event. Moreover, the person can insure against the incurred risk by switching from  $B^*$  to  $B$  by purchasing *the same policy*. Since a single policy does both jobs, the actual change in risk must

be the same in each case. Allais choosers, who perceive a greater risk in the first switch, are committed to paying more for the policy when using it as insurance against the  $A^*$ -to- $A$  risk than when using it as insurance against the  $B^*$ -to- $B$  risk. This difference in “risk premiums” shows that the Allais choosers’ perceptions of risk do not track the actual risks of prospects. Similar things can be said about Ellsberg choosers.

Opponents of expected utility theory may deny that it is appropriate to measure risk by the costs of insuring against it. In the end, the issue will be settled by the development of a convincing method for measuring the *actual* risks involved in prospects. (Ideally, this theory would be augmented by a plausible psychological account of *perceived* risks that explains the common Allais choices.) While there is a well-developed model of risk *aversion* within expected utility theory, this model does not seek to measure risk itself, only an agent’s *attitudes* toward risk. While some progress has been made on the measurement of risk, a great deal remains to be done. It is known that no simple measure (standard deviation, mean absolute deviation, entropy) will do the job. Building on the classic paper Rothschild and Stiglitz (1970), economists have made great strides towards providing a definition of the “riskier than” relation. This work strongly suggests that risk is indeed a separable quantity, and thus that the Allais and Ellsberg choosers are irrational. Still, there is no universally accepted way of measuring the amount of risk that prospects involve. Until such a measure is found the proper interpretation of the Allais and Ellsberg paradoxes is likely to remain controversial, as will expected utility theory itself.

See, also, BAYESIANISM; GAME THEORY.

## References

- Allais, Maurice. "Allais Paradox," in J. Eatwell, M. Millgate, P. Newman, eds., *The New Palgrave: Utility and Probability*, pp. 3-9. New York: Norton, 1990.
- Anderson, Elizabeth. *Value in Ethics and in Economics*. Cambridge, Mass. Harvard University Press, 1993.
- Broome, John. *Weighing Goods*. Oxford. Blackwell Publishers, 1991.
- Ellsberg, Daniel. "Risk, Ambiguity and the Savage Axioms," *Quarterly Journal of Economics* **75** (1961), pp. 643-669.
- Gibbard, Allan and William Harper. "Counterfactuals and Two Kinds of Expected Utility," in C. Hooker, J. Leach, and E. McClennen, eds., *Foundations and Applications of Decision Theory*, pp. 125-62. Dordrecht: Reidel, 1978.
- Hammond, Peter. "Consequentialist Foundations for Expected Utility Theory," *Theory and Decision* **25** (1988), pp. 25-78.
- Jeffrey, Richard. *The Logic of Decision*, 2nd revised edition. Chicago: Chicago University Press, 1983.
- Joyce, James M. *The Foundations of Causal Decision Theory*. New York: Cambridge University Press, 1999.
- Kaplan, Mark. "Decision Theory as Philosophy," *Philosophy of Science* **50** (1983), pp. 549-577.
- Karni, Edi. *Decision Making Under Uncertainty: The Case of State Dependent Preferences*. Boston: Harvard University Press, 1985.
- Kahneman, Daniel and Tversky, Amos. "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica* **47** (1979), pp. 263-291.

- Levi, Isaac. "Consequentialism and Sequential Choice," in M. Bacharach and S. Hurley, eds., *Foundations of Decision Theory*, pp. 92-12. Oxford: Blackwell, 1991.
- McClellenn, Edward. *Rationality and Dynamic Choice: Foundational Explorations*. Cambridge: Cambridge University Press, 1990.
- Ramsey, Frank. "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays*, edited by R. Braithwaite, pp. 156-98. London: Kegan Paul, 1931.
- Raz, Joseph. *The Morality of Freedom*. Oxford: Clarendon Press, 1986.
- Redelmeier, Donald and Shafir, Eldar. "Medical Decision Making in Situations that Offer Multiple Alternatives," *Journal of the American Medical Association* **273** (1995), pp. 302-305.
- Rothschild, Michael and Stiglitz, Joseph. "Increasing Risk: I. A Definition," *Journal of Economic Theory* **2** (1970), pp. 225-243.
- Savage, L. J. *The Foundations of Statistics*. New York: John Wiley and Sons, 1954. 2nd revised edition, New York: Dover Press, 1972.
- Sen, Amartya K. "Choice Functions and Revealed Preference," *The Review of Economic Studies* **38** (1971), pp. 307-317.
- Shafir, Eldar. "Choosing Versus Rejecting: Why Some Options are Both Better and Worse than Others," *Memory and Cognition* **21** (1993), pp. 546-556.
- Shafir, Eldar and Tversky, Amos. "Decision Making," in E. Smith and D. Osherson, eds., *An Invitation to Cognitive Science, Volume 3: Thinking*, (2<sup>nd</sup> edition), pp. 77-100. Cambridge: MIT Press, 1995.
- Skyrms, Brian. *Causal Necessity*. New Haven: Yale University Press, 1980.

Tversky, Amos and Kahneman, Daniel. "Rational Choice and the Framing of Decisions,"

*Journal of Business* (1986), pp. S251-S278.