RECENT WORK

DECISION THEORY

Decision theory seeks to establish general principles of rationality to guide decision makers in situations involving risk or uncertainty. It is a vibrant subject covering many academic disciplines with numerous subspecialties. I cannot hope to do justice to even a small portion of the recent developments in a piece of this length, so I will confine my attention to what I see as the three most fruitful areas of active research. The first concerns the normative status of the *subjective expected utility* (SEU) model that has dominated studies of rational choice for more than fifty years. The second involves *dynamic* decision making in which an agent makes a series of interconnected choices over time. Lastly, I will briefly discuss the renewed importance of decision theory within the theory of games.

The Subjective Expected Utility Model and its Critics

The subjective expected utility (SEU) model remains the primary focus of research within decision theory.[1] It envisions an agent facing a choice among risky or uncertain prospects, called *acts*, whose *outcomes* depend on which of a number of mutually incompatible *states of the world* actually obtains. Her goal is to use what she knows about the world's state to select an act that optimises her chances of securing a satisfying outcome. According to the model, a *rational* agent will always act as if she is *maximising expected utility* relative to some *subjective probability* $\boldsymbol{P}$ that characterises her beliefs and some *utility* $\boldsymbol{u}$ that describes her desires. $\boldsymbol{P}$ is a non-negative, real-valued function, defined over hypotheses about the state of the world, which is *normalised*, so that tautologies are assigned values of one, and *countably additive*, so that the probability of any countable disjunction of pairwise incompatible hypotheses is the sum of the probabilities of its disjuncts. $\boldsymbol{P}$ characterises the agent's beliefs in the weak sense that she regards one hypothesis H as more likely than another H* only if $\boldsymbol{P}$(H) is greater than $\boldsymbol{P}$(H*). $\boldsymbol{u}$ is a real-valued function which, for each act A and state S, assigns a value $\boldsymbol{u}$(A,S) that (subject to the choice of a unit and zero) measures the desirability of the outcome that would ensue were A to be performed when S obtained. To say that the agent behaves 'as if' she maximises expected utility relative to $\boldsymbol{P}$ and $\boldsymbol{u}$ means that, upon being presented with any set of options, she will select one that attains the highest value for the expected utility function $\boldsymbol{EU}(A) = \Sigma_S \boldsymbol{P}(S).\ \boldsymbol{u}(A,S)$.[2] In other words, $\boldsymbol{EU}$ should *represent* her preferences in the sense that she will prefer doing act A to doing act A* only if $\boldsymbol{EU}$(A) > $\boldsymbol{EU}$(A*), and she will be indifferent between the two options only if $\boldsymbol{EU}$(A) = $\boldsymbol{EU}$(A*). An agent whose preferences are representable in this way will automatically maximise expected utility relative to $\boldsymbol{P}$ and $\boldsymbol{u}$ simply by choosing the option she most prefers.

The early work in SEU-theory was taken up with establishing *representation theorems* that enumerate the conditions under which a system of preferences has an expected utility representation.[3] The following four requirements appear, in some version or other, in all such theorems, and each is necessary for rationality on the SEU-model:[4]

> *Transitivity*: If an agent prefers A to B and B to C, then she must also prefer A to C. If she is indifferent between A and B and between B and C, she must also be indifferent between A and C.
>
> *Separability*: An agent's preferences among acts should not depend on what happens in states of the world where those acts bring about identical outcomes.
>
> *Equivalence*: An agent's preferences should not depend on how the options are described to her.
>
> *Continuity*: If B falls between A and C in an agent's preference ranking, there should be a real-number p between zero and one such that B is preferred [dispreferred] to the 'mixed act' in which A is done with probability p [1-p] and C is done with probability 1-p [p].

While satisfaction of these four principles does not alone ensure that an agent will behave like an expected utility maximiser, it is known that they can be augmented with further (non-necessary) constraints that do ensure this. However, no one has yet found a plausible set of conditions on preferences that are both necessary and sufficient for representability. This is perhaps the outstanding unsolved foundational problem in SEU-theory.

Despite its immense influence in the social sciences and philosophy, the SEU-model has recently come under attack on a number of fronts. Psychological research continues to show that thoughtful, well-educated and apparently reasonable people express preferences at odds with basic tenets of the theory, and that they often retain these attitudes even when apprised of their 'errors'. Intransitivities occur in 'preference reversal' experiments where subjects are observed to prefer lottery tickets offering a high probability of modest gains to those offering a low probability of large gains, and yet to set higher 'fair prices' on the latter tickets.[5] Violations of equivalence occur as a result of so-called 'framing effects'. A striking example, due to Daniel Kahneman and Amos Tversky, shows that many people (including physicians!) who are faced with a choice between two forms of medical intervention, A and B, will prefer A when the treatments are described in terms of their impact on mortality rates and B when they are (equivalently) described in terms of their impact on survival rates.[6] Continuity fails in cases where agents see themselves as dealing with incommensurable goods. For instance, most people want more money and prefer living to dying, but when offered a chance to play a (hypothetical) lottery that offers a million dollars with probability p and instant death with probability 1-p many people profess to being unwilling to play when p is any number other than one, even say $1/10^{1000000}$).[7]

People violate the separability principle more frequently and systematically than any of the rest. A well-known example of the phenomena is the *Allais paradox*,[8] in which experimental subjects are presented with two decision problems something like this (where the percentages give the chances of obtaining the various payoffs):

	10%	1%	89%
A	£9,000	£9,000	£9,000
B	£10,000	£0	£9,000

Problem - 1

	10%	1%	89%
A*	£9,000	£9,000	£0
B*	£10,000	£0	£0

Problem - 2

Separability entails that anyone who prefers A to B must prefer A* to B* and conversely. However, many people favour A over B and B* over A*, and these attitudes tend to persist even after it is pointed out that the entries in the rightmost column of each matrix are identical. These results, and others like them, are so pervasive that it is now accepted dogma among psychologists and economists that humans are not expected-utility maximisers.[9]

Troubling as all this is for someone who wants to employ the SEU-model as an account of actual behaviour, proponents of SEU-maximisation usually see their model as having prescriptive force, and thus have largely ignored 'merely empirical' findings. Rational decision makers, they say, *should* be expected utility maximisers even if actual decision makers are not. Nevertheless, it is disquieting that the observed violations of the theory should be so widespread and so persistent among well-educated, successful people. This has led some authors to take a second look at the normative credentials of the SEU-model.

Sequential Choice

Much of the impetus for reevaluation is due to a new appreciation of the subtleties involved in decision making over time. This is relevant to the normative status of the SEU-model because its proponents have traditionally tried to establish transitivity, equivalence, and separability as laws of rationality by presenting arguments which show that, in certain specialised contexts, violators will be disposed to engage in a temporally extended series of transactions that jointly result in their paying costs without accruing benefits. It is precisely these justifications that are being called into question.

To see what is at issue consider the following argument, due to Howard Raiffa,[10] which purports to show that anyone whose preferences violate separability, say by preferring A to B and B* to A* in the Allais example, will throw money away. Begin by supposing that the agent has some preference or other between a sure £9,000 and a lottery that offers a 10/11 chance at £10,000 and a 1/11 chance at £0. Call the first option A** and the second B**, and assume (without loss of generality) that B** is preferred to A**. Now imagine that the agent is presented with the following temporally extended version of the first Allais problem: she begins with the option of choosing B outright or paying a small sum, say a penny, for an arrangement C in which she gets £9,000 with probability 89/100 or a choice between A** and B** with probability 11/100. Since she can select A** if the opportunity arises, it seems that option C is at least as good as getting a sure £9,000, i.e., at least as good as A. So, given that she prefers A to B she will pay the penny to forgo B in favour of C. Unfortunately, she has just paid a penny to forgo B in favour of itself. To see this, notice that C offers an 89% chance of getting £9,000, and an 11% chance of having a choice between A** and B**. In the latter case the agent will certainly choose her preferred option B**. Thus, C is effectively an 89% chance at £9,000, a 10% chance at £10,000, and a 1% chance at £0. This is precisely the option B that she paid to avoid. If the agent had preferred A** to B** we could have made her a similar set of offers involving A* and B* and she would have paid to trade A* for itself. The moral, according to Raiffa, is that agents who exhibit the usual pattern of preferences in the Allais paradox (or who violate separability in any other way) are irrational in the prudential sense because they will willingly incur costs without accruing benefits.

Arguments of this type have recently been receiving a great deal of critical scrutiny. Orthodox decision theory has long held that *extended form* decision problems, in which agents make a series of choices over time, are equivalent to *strategic form* decisions, which involve a one-time, 'up front' choice among contingency plans or *strategies* that stipulate what is to be done in all future eventualities. In Raiffa's example the relevant strategies are: (S_1) choose B; (S_2) choose C and do A** if the opportunity arises; (S_3) choose C and do B** if the opportunity arises. The crucial premise in his argument was the claim that, because S_3 and B end up offering the same outcomes with the same probabilities, the two are identical even though S_3 involves two temporarily separated choices. This clearly assumes the normal form/extended form equivalence.

A number of theorists have begun to doubt this equivalence, and to raise questions about Raiffa's argument (and others like it). The lion's share of the credit for bringing these issues to the forefront of philosophical discussion goes to Frederick Schick, who has shown how much the standard justifications for axioms of SEU-theory owe to non-trivial assumptions about dynamic choice.[11] Economists have become aware of these problems through the efforts of Mark Machina, and Graham Loomes and Robert Sugden (though some of the credit surely goes to earlier work by R. H. Strotz).[12] The upshot has been an explosion of work on the nature of decision making over time.

Three broad positions have been staked out: Peter Hammond's *consequentialism* is a defence of both SEU-theory and the normal/strategic form equivalence from general principles about the proper relation of actions to their consequences;[13] Edward McClennen's *resolute choice* account, which has affinities to views expressed by Mark Machina and David Gauthier, sees dynamic decision making as a matter of making sensible plans and sticking to them;[14] and, lastly, the doctrine of *sophisticated choice* asks agents to treat their future choices as they would treat any other, contingent event that might affect their happiness.[15] We will consider these views in turn.

Hammond's consequentialism represents the most serious attempt to date at saying what it means for actions to be 'valued by their consequences'. His view has two central components:

> *Dynamic Consistency*: A strategy for making a sequence of decisions can only be rational if it remains rational at each stage in the sequence when all previous decisions are ignored (save for their effects on the likelihoods of future outcomes).
>
> *Consequentialism*: The right choice in any dynamic decision situation should depend only on the consequences available and their likelihoods.

The first principle says that a rational strategy must recognise that its agent will always be 'looking down the road' by ignoring risks already borne, opportunities already missed, and costs already sunk. This makes S_2 an irrational strategy because S_2 asks the agent to choose A** rather than B** at the second choice point, which is something she would not do if given a straight choice between these two options. The second principle says that a rational agent's choice in any sequential decision problem should not depend on the structure of the 'decision tree' she faces. For example, two courses of action that differ only in the order in which choices are made should be regarded as equally desirable. This is the normal/extensive form equivalence in consequentialist garb. Hammond shows that conforming to dynamic consistency and consequentialism leads to preferences that obey transitivity, equivalence and separability. The basic principles of SEU-theory (save continuity) are thus seen to be natural consequences of the notion that actions are valued by their consequences.

Hammond's work is controversial. Edward McClennen, who rejects transitivity and independence, has questioned dynamic consistency on the ground that rational agents ought to execute strategies *resolutely* by, "continuing to implement the plan *initially* settled on, so as to ensure the sequences of choices thus made serves to access the prospect *initially* judged to be most acceptable"[16] (emphasis added). McClennen denies, for example, that someone who prefers B** to A** must choose B** over A** in Raiffa's example. Rather, he maintains, a person can choose S_2 as a way of attaining the best overall result A, and, having chosen it, she can rationally perform A** at the second choice point. The motivating idea seems to be that whatever facts make it rational to choose a plan also make it rational to act in accordance with the plan, and this is supposed to hold even in cases where these acts would *not* be rational taken by themselves. The main challenge for such a view is to explain why an agent should act on plans that, by her current lights, promise less than optimal results merely because these plans looked good at some earlier time. What one seems to have here is a tyranny of present selves over future selves. McClennen tries to make this palatable by suggesting that the adoption of an 'up front' strategy in a sequential decision problem is a kind of bargain that an agent strikes with the person she will be at the time of future choices. This seems implausible, however, given that the agent's future self does not exist at the time the agreement is made, and that there is no enforcement mechanism to ensure that the 'bargain' will be honoured.[17]

Mark Machina accepts something like resolute choice, but he offers a different criticism of Hammond's position. While he endorses dynamic consistency, he suggests that nonseparable preferences can sometimes be rational and that consequentialism is inappropriate in such instances. The example he offers is one in which a mother would rather have some good awarded to one of her two children by a fair coin flip than have it go to the oldest by fiat, this despite the fact that she sees no advantage in the good going to the younger child. The mother seems to violate separability here because she prefers (youngest if heads, oldest if tails) to (oldest if heads, oldest if tails) without also preferring youngest to oldest in the event of heads. While Machina grants that it would not be rational for the mother to maintain these preferences along with a consequentialist approach to dynamic decision making, he sees this as telling against consequentialism rather than nonseparable preferences. The force of his conclusion, of course, rests on the plausibility of the claim that the mother's preferences are both nonseparable and rational. No one would dispute that they are rational, but it does seem debatable that they are nonseparable. John Broome, for example, has pointed out that the apparent nonseparability of the mother's preferences vanishes once prospects are redescribed as (youngest gets the good fairly if heads, oldest gets the good fairly if tails) and (oldest gets the good unfairly if heads, oldest gets the good unfairly if tails).[18] This seems like the right thing to say about the case at hand. Still, the general strategy of avoiding violations of the SEU-model by redescribing outcomes is fraught with pitfalls, and it is not clear that all cases of seemingly rational but nonseparable preference can be dispatched so easily.[19]

Isaac Levi also rejects Hammond's position,[20] arguing for a weaker form of consequentialism that does not presuppose the strategic/extensive form equivalence. As Levi notes, Hammond's view does not allow an agent to regard his future states of belief or desire in quite the same light as he regards other states of the world. The point is nicely brought out in the 'potential addict' problem. Here an ex-smoker is wondering whether to have a cigarette. Having just one would be a good thing, but having more would be awful. If we consider just the normal form version of his decision—in which the relevant acts are to have just one cigarette, to have more than one, or to abstain—it seems clear that the man should have a cigarette. The rub is that he is sure to become readdicted if he does this, i.e., he knows that having even one cigarette will so drastically alter his preferences (or opinions) that he will be incapable of stopping himself from having more. Abstinence is obviously the best policy here.[21] Hammond's view does not yield this answer however. Indeed, it is not clear that it yields any answer because changes that occur in the agent's preferences over time simply do not get factored into the consequentialist framework.

Hammond wants to suggest that the potential addict problem is not a genuine difficulty for his approach, because "the potential addict is really two (potential) persons, before and after the addiction, and the decision problem must be analyzed as a 'game' between two 'rational players' ".[22] This is not terribly plausible, and even if it were it would not be of much help. It suggests that consequentialist thinking is only appropriate for situations where an agent's beliefs and desires at each future choice point are identical to his initial beliefs and desires conditional on being at that point. This is a fairly significant restriction since it rules out all cases in which an agent can anticipate her future self yielding to a 'weakness of the will'. Rather than accept this constraint, Levi defends a weakened form of consequentialism that lets agents make 'sophisticated' choices by treating their future beliefs, desires, and acts as objects of uncertainty on a par with states of the world. While this leads him to reject the standard dynamic arguments for transitivity, he believes that separability follows from his weakened consequentialism via an argument due to Teddy Seidenfeld.[23]

The ultimate resolution of these issues about dynamic choice will have a major impact on the future of the SEU-model.[24] Scepticism about transitivity and separability has already led a number of authors to propose alternatives to SEU-maximisation. Graham Loomes and Robert Sugden, and Peter Fishburn have proposed a skew-symmetric bilinear model that seeks to represent rational preference using a relation:

$$\text{A is preferred to B if } \mathbf{SSB}(A,B) = \Sigma_S \mathbf{P}(S).\mathbf{F}_S(A,B) > 0$$

where $\mathbf{F}_S$ is a skew-symmetric bilinear function[25] that measures the extent to which the agent would *regret* having A rather than B in state S.[26] This representation allows intransitive preferences. However, it also lets agents be indifferent between choiceworthy and non-choiceworthy acts, which makes it unclear why preferences should serve as guides to action in the first place.

Mark Machina rejects the separability axiom and proposes a *generalised expected utility analysis* that replaces standard ***EU*** representations, which are linear functions of probabilities, with functionals of the form (for S1, S2,. . . states of the world and A an act):

$$\mathbf{GEU}(A) = \mathbf{F}(\mathbf{P}(S1),\mathbf{u}(A,S1);\ \mathbf{P}(S2),\mathbf{u}(A,S2);\ldots;\ \mathbf{P}(Sn),\mathbf{u}(A,Sn))$$

where **F** may be almost any differentiable mapping that is strictly increasing in the utilities. Such a **GEU** can be approximated arbitrarily closely by an ordinary expected utility function when the basic outcomes are 'sufficiently close' in utility. This has the effect of making rational preference 'locally' separable so that violations of SEU-theory only occur when utility differences between outcomes are large. This dovetails nicely with the experimental evidence, which shows that people tend to violate separability in Allais-type cases only when the payoffs are far apart. However, it does presuppose that nonseparable preference can be rational. Other non-expected utility models have been proposed by S. Chew, J. Hey, I. Levi, U. Segal, and M. Yarri.[27]

Despite this host of competitors, and the empirical and foundational issues that spawned them, the SEU-model is far from being dead. Indeed, some of its greatest successes have come during the last few years. Philosophers have found ways to consistently introduce causal notions into the theory, and representation theorems for these *causal decision theories* have been proved.[28] Richard Jeffrey's maxim of *ratifiability*, which asks agents to choose acts that maximise expected utility on the hypothesis that they are decided upon (as opposed to the hypotheses that they are performed), has gone from an inventive, but unsuccessful attempt to avoid talk of causality to an important tool within causal decision theory itself.[29] John Broome[30] has recently argued that the constraints that the SEU-model imposes on preferences are also appropriate for goodness *per se*. Specifically, he maintains both that it makes sense to speak about 'the good' itself, and that the structure of the good is correctly described by the axioms of SEU-theory. Finally, and perhaps most importantly, the SEU-theory is assuming an increasingly important role in the theory of games, where it has historically been something of a bit player. It is to this development that we now turn.

Decision Theory and Game Theory

While decision theory is concerned with the actions of individuals, game theory aims to describe the behavior of groups of agents in choice situations where the outcome each 'player' receives depends on what others do.[31] The SEU-model has always been part of game theory, in some ways its foundation, but until very recently game theorists have not had to pay much attention to its details. In a famous passage in *Theory of Games and Economic Behavior*, John von Neumann and Oscar Morgenstern argued that rational players who know (i) all there is to know about the structure of the game they are playing, (ii) all there is to know about the opinions and motivations of other players, (iii) that everyone in the game is rational, (iv) that everyone knows (i)–(iii), (v) that everyone knows (i)–(iv), and so on,[32] ought to be able to deduce the optimal strategy for every player (and should recognise that everyone else can do likewise). In such circumstances, the total set of strategies chosen by all players will be in equilibrium: each action will be a 'best reply' to all the rest in the sense that it will maximise its players' expected utility conditional on what he expects the others to do. Here is a simple illustration:

	C1	C2
R1	2,1	0,0
R2	0,0	0,0

Player ROW chooses beteen strategies R1 and R2 and player COLUMN between C1 and C2. Their joint action fixes a utility payoff for each player, for example R1 and C1 yield ROW a utility of 2 and COLUMN a utility of 1. This game has 'pure' equilibria at (R1,C1) and (R2,C2).

Von Neumann and Morgenstern showed that every zero-sum game has an equilibrium, and John Nash extended this result to all games on the supposition that players could choose 'mixed acts' by letting their pure acts be determined by randomising devices. Ever since, it has remained the 'central dogma' of game theory that rational players will always choose their end of a Nash equilibrium. Moreover, since von Neumann and Morgenstern's argument presupposes almost nothing about the nature of the deliberative processes that lead players to choose specific equilibria, game theorists have been able to ignore these processes almost entirely and to focus on the equilibria themselves.

The one fly in the ointment was the existence of multiple, non-equivalent equilibria. Many games contain equilibria that seem 'sensible', like (R1,C1) above, and others that do not, like (R2,C2). It would be desirable to have a well-motivated way of ruling out the latter. The 'refinement programme', initiated by Reinhard Selten in 1965, took important steps toward this end by defining new, more restrictive solution concepts that leave fewer equilibria open for rational agents to (collectively) choose.[33] In the game above, for example, Selten would rule out (R2,C2) on the grounds that it lacks a kind of stability found in (R1,C1). Specifically, unless ROW is absolutely certain that COLUMN will be able to execute her choice of C2 then ROW will surely not want to play R2. This is not true of R1 and C1. In the jargon, (R1,C1) is a *perfect* equilibrium while (R2,C2) is not. More sophisticated solution concepts have been developed to rule out other problem cases.

By any measure Selten's programme has been a smashing success (as indicated by his receipt of the most recent Nobel Prize in economics, along with Nash and Harsanyi). However, it has not entirely solved the problem since, for any refinement of the equilibrium concept one introduces, there remain problem cases containing non-equivalent equilibria. This has led some game theorists to conclude that one must delve into the details of the deliberative process in order to make any real headway. The bellwether for this movement was Ken Binmore's pair of articles 'Modelling Rational Players I and II',[34] which invited game-theorists to search for 'educative procedures' that describe the inferential processes rational agents might employ in reasoning themselves into equilibrium choices.

A great deal of progress has been made on this front, much of it by philosophers. Brian Skyrms has developed a sophisticated account of rational deliberation for SEU-maximisers that entails the von Neumann/Morgenstern equilibrium requirement, and which shows that distinctions among different sorts of equilibria can be explicated in terms of differences in the way an agent would arrive at them.[35] The concept of a 'perfect' equilibrium, for example, is explained not in terms of players being unable to execute their acts, but in terms of the wide range of initial beliefs that would lead a rational deliberator to choose the equilibrium in question. William Harper has taken up Binmore's challenge as well by developing a deliberative procedure that solves some important classes of games using causal decision theory supplemented with an iterated use of Jeffrey's maxim of ratifiability.[36] Harper's method is particularly useful in solving sequential games in which players are able to 'communicate' their intentions through their actions. Even though this work is still in its infancy, I fully expect it to have lasting significance. There is no doubt that it will serve as an interesting and important arena for philosophical disputes about rationality for many years to come.

Notes

1. The classic source is Leonard Savage's *The Foundations of Statistics*, 2nd revised edition (Dover, 1972). A less technical, but more philosophically satisfying treatment can be found in Richard Jeffrey, *The Logic of Decision*, 2nd edition (University of Chicago Press, 1983). David M. Kreps, *Notes on the Theory of Choice*, (Westview Press, 1988) is a splendid recent addition to the literature. The mathematically sophisticated reader will find a fine overview of SEU-theory in Peter Fishburn, 'Subjective Expected Utility: A Review of Normative Theories', *Theory and Decision*, vol. xiii (1981), pp. 139–99. Two useful collections of essays are Michael Bacharach and Susan Hurley (eds.), *Foundations of Decision Theory* (Basil Blackwell, 1991); and Peter Gardenfors and Nils-Eric Sahlin (eds.), *Decision, Probability and Utility* (Cambridge University Press, 1988). Finally, two 'handbooks' that deserve a place on every serious decision theorist's bookshelf are John Eatwell, Murray Milgate, and Peter Newman (eds.), *Utility and Probability* (W. W. Norton & Company, 1990); and S. Barbera, P. J. Hammond, and C. Seidl (eds.), *Handbook of Utility Theory* (Kluwer, 1995).
2. Proponents of the SEU-model do not imply that the agent actually makes her choices by *calculating* expected utilities. Her internal decision-making process is not at issue. Rather, the model aims to give an *external* account of what a rational decision maker's choices will look like once they are made: i.e., it is meant as a *substantive* rather than a *procedural* account of rationality. For a useful discussion of this point and related issues see Herbert Simon, 'Rationality as Process and as the Product of Thought', *American Economic Review*, vol. lxviii (1978), pp. 1–16.
3. See Fishburn, *op. cit.*, for a useful discussion of the history.
4. An excellent non-technical discussion of these principles can be found in John Broome, *Weighing Goods* (Basil Blackwell, 1991).
5. See, for example, David Grether and Charles Plott, 'Economic Theory of Choice and the Preference Reversal Phenomena', *American Economic Review*, vol. lxix (1979), pp. 623–638.
6. Amos Tversky and Daniel Kahneman, 'The Framing of Decisions and the Psychology of Choice', *Science*, vol. ccxi, pp. 453–458.
7. For a philosophical defence of the preferences described see David Velleman, 'The Story of Rational Action', *Philosophical Topics*, vol. xxi (1993), pp. 229–254. Relevant psychological results can be found in R. Thaler, 'Mental Accounting and Consumer Choice', *Marketing Science*, vol. iv (1985), pp. 199–214.
8. The paradox was first introduced in Maurice Allais, 'Le Comportement de l'homme rationnel devant le risque: critique des postulates ex axiomes de l'école Americaine', *Econometrica*, vol. xxi (1953), pp. 503–546. Useful articles on the paradox can be found in M. Allais and O. Hagen (eds.), *Expected Utility Hypotheses and the Allais Paradox* (Reidel, 1979).

9. Useful overviews of the psychological results can be found in Daniel Kahneman and Amos Tversky, 'Prospect Theory: An Analysis of Decision Under Risk', *Econometrica*, vol. xlvii (1979), pp. 263–291; and Mark Machina, 'Generalized Expected Utility Analysis and the Nature of Observed Violations of the Independence Axiom', in B. Stigum and F. Wenstop (eds.), *Foundations of Utility and Risk Theory with Applications* (Reidel, 1983), pp. 117–136.
10. Howard Raiffa, *Decision Analysis: Introductory Lectures on Choice Under Uncertainty* (Addison Wesley, 1968), pp. 82–83.
11. Frederick Schick, 'Dutch Books and Money Pumps', *Journal of Philosophy*, vol. lxxxiii (1986), pp. 112–118.
12. Mark Machina, ' 'Rational' Decision Making versus 'Rational' Decision Modeling? A Review of Allais and Hagen (eds.), *Expected Utility and the Allais Paradox*', *Journal of Mathematical Psychology*, vol. xxiv (1981), pp. 163–175; Graham Loomes and Robert Sugden, 'Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty', *Economic Journal*, vol. xcii (1982), pp. 805–824; R. H. Strotz, 'Myopia and Inconsistency in Dynamic Utility Maximization', *Review of Economic Studies*, vol. xxiii (1955), pp. 165–180.
13. See Peter Hammond, 'Consequentialist Foundations for Expected Utility', *Theory and Decision*, vol. xxv (1988), pp. 25–78, and 'Consistent Plans, Consequentialism, and Expected Utility', *Econometrica*, vol. lvii (1989), pp. 1445–1449.
14. See Edward McClennen, *Rationality and Dynamic Choice* (Cambridge University Press, 1990); Mark Machina, 'Dynamic Consistency and Non-Expected Utility', in Bacharach and Hurley, *op. cit.*, pp. 39–91; and David Gauthier, 'In the Neighborhood of the Newcomb Predictor (Reflections on Rationality)', *Proceedings of the Aristotelean Society*, vol. lxxxix (1988–89), pp. 179–94.
15. A defence of sophisticated choice can be found in Isaac Levi, 'Consequentialism and Sequential Choice', in Bacharach and Hurley, *op. cit.*, pp. 92–122.
16. McClennen, *op. cit.*, p. 159.
17. For a useful critique of McClennen's view see Daniel Farrell, 'Utility Maximizing Intentions and the Theory of Rational Choice', *Philosophical Topics*, vol. xxi (1993), pp. 53–78.
18. Broome, *op. cit.*, pp. 107–115.
19. For useful discussion of this 'redescription strategy', see Velleman, *op. cit.*, 235–241; and Philip Pettit, 'Decision Theory and Folk Psychology', in Bacharach and Hurley, *op. cit.*, pp. 163–166.
20. Levi, *op. cit.*, pp. 106–114.
21. We are assuming here that the agent does not have the option of, say, destroying all but one cigarette before he smokes. This would still be an interesting problem, but one of a different kind. For relevant discussion see Jon Elster, *Ulysses and the Sirens* (Cambridge University Press, 1979).
22. Hammond, 'Consequentialist Foundations', pp. 36.
23. Teddy Seidenfeld, 'Decision Theory Without 'Independence' and Without 'Ordering': What is the Difference', *Economics and Philosophy*, vol. iv (1988), pp. 267–290.
24. It will also have a major impact on the theory of subjective probability since Bayesian arguments for conditionalisation often involve dynamic decision making. There is a vast literature here. Some useful references are: Paul Teller, 'Conditionalization and Observation', *Synthese*, vol. xxvi (1973), pp. 218–258; Brad Armendt, 'Is There a Dutch Book Argument for Probability Kinematics', *Philosophy of Science*, vol. xlvii (1980), pp. 583–588; Bas van Fraassen, 'Belief and the Will', *Journal of Philosophy*, vol. lxxxi (1984), pp. 235–256; Brian Skyrms, 'A Mistake in

Dynamic Coherence Arguments?', *Philosophy of Science*, vol. lx (1993), pp. 320–328; Patrick Maher, *Betting on Theories* (Cambridge University Press, 1993).

25. That is, $\mathbf{F}_S$ is a continuous, real-valued function such that $\mathbf{F}_S(A,A) = 0$, $\mathbf{F}_S(A,B) = -\mathbf{F}_S(B,A)$, and for each real number p between zero and one $F_S([pA + (1-p) A^*],B) = p\mathbf{F}_S(A,B) + (1-p)\mathbf{F}_S(A^*,B)$ where $[pA + (1-p)A^*]$ is the 'mixed act' of doing A with probability p and A* with probability 1–p.
26. Graham Loomes and Robert Sugden, *op. cit*,; Peter Fishburn, 'SSB Utility Theory: an Economic Perspective', *Mathematical Social Sciences*, vol. viii (1984), pp. 63–94.
27. S. Chew, 'A Generalization of the Quasi-linear Mean With Applications of the Measurement of Inequality and Decision Theory Resolving the Allais Paradox', *Econometrica*, vol. li (1983), pp. 1065–1092; J. Hey, 'The Economics of Optimism and Pessimism: a Definition and some Applications', *Kyklos*, vol. xxxvii (1984), pp. 181–205; I. Levi, *Hard Choices* (Cambridge University Press, 1986); U. Segal, 'The Ellsberg Paradox and Risk Aversion: an Anticipated Utility Approach', *International Economics Review*, vol. xxviii (1987), pp. 175–202; M. Yarri, 'The Dual Theory of Choice Under Risk', *Econometrica*, vol. lv (1987), pp. 95–115.
28. Causal decision theory asks agents to choose acts that have the best causal consequences. It is contrasted with *evidential decision theory*, which has them choose acts that are indicative or symptomatic of desirable results. For a useful recent discussion of work in this area see: Allan Gibbard and Jim Joyce, 'Causal Decision Theory', in Barbera, Hammond, and Seidl, *op. cit.*
29. Jeffrey, *The Logic of Decision*, pp. 15–20. A useful critical discussion of ratifiability can be found in W. Rabinowicz, 'Ratifiability and Stability', in Gardenfors and Sahlin, *op. cit.*, pp. 406–425.
30. Broome, *op. cit.*
31. The classic sources are: John von Neumann and Oscar Morgenstern, *Theory of Games and Economic Behavior*, 2nd edition (Princeton University Press, 1947), pp. 148; and Duncan Luce and Howard Raiffa *Games and Decisions* (Wiley, 1957). An excellent recent introduction (just right for a philosopher coming to the subject anew) is David Kreps *Game Theory and Economic Modeling* (Clarendon Press, 1990).
32. These are substantive assumptions. There is an active research programme that is seeking to see how much they can be relaxed without destroying the theory altogether. Philosophers interested in this issue may want to look at: Martin Hollis and Robert Sugden, 'Rationality in Action', *Mind*, vol. cii (1993), pp. 1–33; and Cristina Bicchieri, 'Paradoxes of Rationality', *Midwest Studies in Philosophy*, vol. xv (1990), pp. 65–79.
33. A good overview of this work can be found in John Harsanyi and Reinhard Selten, *A General Theory of Equilibrium Selection in Games* (MIT Press, 1988).
34. Kenneth Binmore, 'Modelling Rational Players I', *Economics and Philosophy*, vol. iii (1987), pp. 179–214; 'Modelling Rational Players II', *Economics and Philosophy*, vol. iv (1988), pp. 9–55.
35. Brian Skyrms, *The Dynamics of Rational Deliberation* (Harvard University Press, 1990).
36. William Harper, 'Mixed Strategies and Ratifiability in Causal Decision Theory', *Erkenntnis* vol. xxiv (1986), pp. 25–36; 'Ratifiability and Refinements', in Bacharach and Hurley, *op cit.*, pp. 262–293.

THE UNIVERSITY OF MICHIGAN JAMES JOYCE