MIDWEST STUDIES IN PHILOSOPHY, XV (1990)

Ratifiability and the Logic of Decision[1]

BRIAN SKYRMS

Happiness is achieved by prudence: prudence is found in right actions: a right action is one that, once performed, has a probable justification.

—Arcesilaus

1. INTRODUCTION

Richard Jeffrey introduced an equilibrium concept into the theory of individual rational decision in the second edition of *The Logic of Decision*. He called the concept *ratifiability*.[2] The system of the book originally was motivated by a desire to deal with cases in which states are not independent of acts and to do so without causally loaded concepts. Probabilities are defined on a large boolean algebra, whose elements are taken to be propositions and whose operations are to be taken as truth functions. Acts are construed as propositions in this space that can be directly "made true" by the decision-maker. The inclusion of acts in the boolean algebra over which probabilities are defined is an innovation which may provoke varying reactions.[3] However, one may argue that this feature makes the system attractive for dealing with sequential decision problems. In such problems, the choice of an option may change its status over time from consequence to act to part of the state of the world, and each change goes with an appropriate updating of subjective probability.

But examples like Newcomb's problem, where acts are evidentially relevant but not causally relevant to preexisting states of the world, convinced Jeffrey that choiceworthiness does not always go by the evidential conditional expected utility of his system. Savage's theory, with a suitable choice of states outside the causal influence of the acts, gives the right answers.[4]

Jeffrey introduced the concept of ratifiability to deal with these problematic cases within the framework of his system. Informally, a ratifiable act is one that is optimal if chosen. The erroneous choices were to be eliminated by the requirement that the act chosen be ratifiable. However, ratifiability was not given any precise definition within Jeffrey's system, and the discussion of its sensitivity to causal considerations has been inconclusive.

I will discuss here a precise definition of ratifiability within a Jeffrey-type framework supplemented with Savage-type distinctions. His definition is not new. It has, in fact, been discussed by both economists and philosophers. Given the definition, a certain version of the doctrine of ratificationism[5] is clearly correct. Pursuit of a ratifiable act cannot lead to an act which is not choiceworthy. However, another version of the doctrine of ratificationism—roughly the hypothesis that causal and evidential decision theory must agree at the moment of action—is much more problematic.

The plan of this essay is as follows: Section 2 sets up a framework in which Jeffrey-type and Savage-type expected desirabilities can be compared and gives a definition of ratifiability within this framework. Section 3 shows one sense in which the doctrine of ratificationism is clearly correct and section 4 discusses a different sense in which it is problematic. Section 5 notes the importance of ratifiability in the theory of games. Section 6 discusses the status of ratifiability as a principle of rationality. Section 7 comments on the significance of ratifiability for decision theory. Finally, section 8 serves as a brief guide to the literature.

2. DEFINITION OF RATIFIABILITY IN A JEFFREY-SAVAGE FRAMEWORK

Before we can discuss the relationship between Savage and Jeffrey decision rules, we need a common framework in which they both can operate. In Jeffrey's framework, probability and value are defined on a common boolean algebra of propositions. Savage's framework distinguishes acts, states, and consequences with probabilities defined on the space of states, and acts being functions from states to consequences. In the application of Jeffrey's framework to a given decision problem, however, one needs to introduce one of the distinctions that is built into the Savage framework. One must identify a partition of *act*-propositions, each of which is "within the agent's power to make true if he pleases,"[6] that represents the relevant acts for the decision problem under consideration. Identification of the relevant act-propositions requires causal judgment. We will assume here a further exercise of causal judgment in identifying a partition of propositions which are surrogates for Savage states. The states are outside the influence of the decision-maker, and they, together with the acts, capture the causal conditions relevant to the payoffs.

Let us assume, for simplicity, a decision problem with a finite number of acts and states such that the acts do not causally influence the states, and the states together with the acts jointly determine the value of payoffs. Acts and states can now be thought of as partitions: $\{A_i\}$; $\{K_j\}$ respectively, on a Jeffrey space. We will assume that on the elements of the common refinement of these partitions, Jeffrey value and Savage utility coincide. This last assumption is made purely for convenience of exposition. Where it does not hold, a slightly more complicated form of causal decision theory is appropriate.[7]

Jeffrey expected value is defined for every proposition, P, and any partition, $\{Q_i\}$ as:

$$V(P) = \Sigma_i \operatorname{pr}(Q_i|P)\, V(Q_i \,\&\, P).$$

Savage expected utility is defined here for acts as:

$$U(A) = \Sigma_j \operatorname{pr}(K_j)\, V(A \,\&\, K_j).$$

For any proposition, P, we have Savage expected utility conditional on that proposition as:

$$U(A|P) = \Sigma_j \operatorname{pr}(K_j|P)\, V(A \,\&\, K_j).$$

This is what the Savage expected utility would be if one conditioned on P to get new probabilities of the states. In this connection, it is worth noting that here the Jeffrey expected value of an act is just the Savage expected utility of that act conditional on itself:

$$(E)\ V(A) = U(A|A).$$

There is only one formal definition of ratifiability that makes sense in this framework:

Def. (R): A_i is *ratifiable* iff $U(A_i|A_i) \geq U(A_j|A_i)$ for all j.

By *(E)*, one can just as well say that A_i is ratifiable just in case its Jeffrey expected value is at least as great as the Savage expected utility conditional on it of each of its competitors:

(R') A_i is *ratifiable* iff $V(A_i) \geq U(A_j|A_i)$ for all i.

Jeffrey expected value is, in this sense, the figure of merit for the ratifiable act—but not for its contrast class.

Investigation of the concept of ratifiability requires use of elements from both the Jeffrey and the Savage frameworks. Ratifiability cannot be defined in a Savage framework because the relevant conditional probabilities, of states conditional on acts, do not exist. Ratifiability can be defined in the Jeffrey framework only when Savage's distinction between acts and states has been introduced.

3. RATIFIABILITY AND CHOICEWORTHINESS (YES)

Ratificationists[8] agree that an act which maximizes Savage expected utility is choiceworthy, while—in the tricky cases—one which maximizes Jeffrey expected value may not be. The injunction: "Choose a ratifiable act!" is supposed to guide decision-makers in operating within a Jeffrey framework to a choiceworthy act. Therefore, ratificationists cannot support any definition of ratifiability in substantive disagreement with that given in the previous section. On that definition, a ratifiable act is just one which maximizes Savage expected utility conditional on the hypothesis that it is carried out. In what sense can we show that pursuit of a ratifiable act cannot lead to a decision which is not choiceworthy?

We will consider three increasingly general models of deliberation. In the simplest model of deliberation the decision-maker chooses a ratifiable act, executes it, and conditions on the proposition that the act is done—and is then in a belief state in which Savage expected utility is maximized. In this sort of a model, prior ratifiability by definition coincides with posterior choiceworthiness, where posterior choiceworthiness is measured by Savage expected utility at the moment of truth.

In a slightly more sophisticated model of deliberation, decision-makers' beliefs change by probability kinematics on the acts (that is, the probabilities of the states conditional on the acts remain constant) until one act is done and gets probability 1. This model does not differ from the previous one in the relation of the starting to stopping points, but only on the description of the road in between. The posterior probabilities are gotten from the prior probabilities by conditioning on the act chosen. So again, by definition, choice of an act which is ratifiable *a priori* will lead to an act which *a posteriori* maximizes Savage expected utility.

But deliberation may be a much more complicated process than the simple model just described. Suppose deliberation is a process on the temporal interval [0,1], with the decision-maker being initially unsure at t_0 of which act to perform, and finally arriving at probability 1 for some act at t_1. Rather than assuming that the probabilities of the states conditional on the acts remain constant as in belief change by conditioning or probability kinematics on the acts, let us consider *any* deliberational process where the probabilities of the states conditional on the acts (where defined) change continuously with respect to time. In the large class of deliberational models satisfying this assumption many bizarre things can happen. Certainly it is possible that an initially ratifiable act may cease to be ratifiable during deliberation. But even in this model, there is still one precise sense in which we can easily show that pursuit of a ratifiable act cannot lead one astray. Conditioning and kinematics on the acts are special cases.

Theorem: If an act chosen, A^*, which gets probability 1 at time t_1, remains ratifiable for some stretch of time $[x,1]$ with $0 \leq x < 1$ up to the moment of truth and if the deliberational process makes $\mathrm{pr}(K_j|A_i)$ change continuously with time on [0,1], then A^* maximizes Savage expected utility at the moment of truth (t_1).

Proof: Suppose not. Then there is an act, A', such that at t_1 $EU(A') > EU(A^*)$. But at t_1, $EU(A_i) = EU(A_i|A^*)$ because $\mathrm{pr}(A^*) = 1$. So at t_1 $EU(A'|A^*) > EU(A^*|A^*)$. By hypothesis for all j, $\mathrm{pr}(K_j|A^*)$ changes continuously with time on [0,1]. For all i, $EU(A_i|A^*)$ is a continuous function of the probabilities: pr $(K_j|A^*)$. So $EU(A'|A^*)$ and $EU(A^*|A^*)$ vary continuously with time and so does their difference. By continuity throughout some neighborhood of $t = 1$, $EU(A'|A^*) > EU(A^*|A^*)$ contradicting the hypothesis that A^* is ratifiable throughout $[x,1)$.

4. RATIFIABILITY AND CHOICEWORTHINESS (NO)

Might it be the case that Jeffrey expected value and Savage expected utility coincide at the moment of truth? This is the leading idea of a number of discussions of the impact of causal counterexamples on Jeffrey's system. The question as it stands is ill-defined. The problem is not with the expected value of the act chosen – under the assumptions in force its Jeffrey expected value is indeed equal to its Savage expected utility[9] – but with the Jeffrey expected value of its competitors. This is not well defined because the relevant conditional probabilities are on conditions which at t_1 have probability 0.[10] (Thus an attempt to substitute V for U in definition R would produce nonsense.)

The next best thing is to compare the Savage expected utilities at the moment of truth with the limit of the Jeffrey expected values as the decision-maker approaches the moment of truth along an orbit of deliberation. The question is then meaningful only with respect to some model of deliberation. In certain special cases, a plausible model can indeed give the result that Jeffrey and Savage agree in the limit.[11] But can these special cases be extended for some reasonable model of deliberation to a general theorem?

Let us consider models of deliberation on the unit interval, as in the previous section, where the acts have probabilities unequal to 0 or 1 on t in [0,1], and one of the acts gets probability 1 at $t = 1$. The general case would require deliberation starting at any coherent prior probability in this class to end with Jeffrey and Savage in agreement in the limit. It seems reasonable to assume continuity of change of the probabilities of the states conditional on the acts throughout deliberation, as was also done in the previous section, in order to rule out imposition of agreement between Jeffrey and Savage by a kind of deliberational miracle. And it seems reasonable to require that deliberation be *coherent*. That is to say that it should not be the case that starting from a coherent prior the model of deliberation postulated leaves one open to a dutch book: a finite number of bets which one judges fair or favorable such that one suffers a net loss for every possible outcome. Under these conditions, we can show that no such deliberational model can exist.

Theorem: There is no coherent model of deliberation under which the probabilities of the states conditional on the acts change continuously with time, such that for any coherent prior Jeffrey and Savage coincide in the limit.

Proof: Consider the decision problem with the following payoffs and initial probabilities:

	Payoff		*Probabilities*	
	State 1	State 2	State 1	State 2
Act 1	1	0	.9	0
Act 2	0	2	0	.1

This can be thought of as a version of "The Nice Demon." A nice demon has predicted whether the decision-maker will chose 1 or 2 and has arranged the state so that the decision-maker will be rewarded if the nice demon's prediction is correct. According to the decision-maker's initial probabilities, there is .9 probability that the demon predicted Act 1 and arranged for State 1 to obtain; .1 probability that Act 2 was predicted and State 2 obtains. The decision-maker believes that the demon will be correct with probability 1. (To recast the example as a game, one may think of the decision-maker as playing a pure coordination game with the nice demon, and using "best response" reasoning.)

Coherence requires that the initial zero probabilities not be raised. Otherwise the decision-maker is open to a trivial dutch book. In our example, at time t_0 a cunning bettor buys from the decision-maker a bet which pays $1 to the bettor if Act 1 and State 2 or Act 2 and State 1 obtains; nothing otherwise, for its fair price of exactly nothing. At time t_1 she sells back the bet for a price equal to the current probability of Act 1 and State 2 or Act 2 and State 1, making the dutch book.

So in a coherent model of deliberation, for all $t<1$:
pr(State 1|Act 1) = pr(State 2|Act 2) = 1

because zeros are not raised and no act gets probability 1 until $t = 1$. At $t = 1$, one of these conditional probabilities goes undefined, but the other gets probability 1 by continuity. Jeffrey values of Acts 1 and 2 must be 1 and 2 respectively throughout deliberation so these must be the limiting values at $t = 1$. At $t = 1$, the decision-maker has decided which act to do, and that act gets probability one. If this act is Act 1 then at $t = 1$ pr(State 1 and Act 1) = 1, because at $t = 1$ pr (State 1|Act 1) = 1. By similar reasoning, the other alternative is that at $t = 1$ pr (State 2 and Act 2) = 1. In neither case does Savage expected utility coincide with Jeffrey expected value. If Act 1 is chosen, the Savage expected utility of Act 1 is 1 and of Act 2 is 0; if Act 2 is chosen, the Savage expected utility of Act 1 is 0 and of Act 2 is 2.[12]

5. RATIFIABILITY AND EQUILIBRIUM

Ratifiability is a kind of equilibrium concept for rational decision. As such, it takes on special significance in the context of strategic interaction among rational decision-makers—that is, in the sort of problem customarily treated by the theory of games. In fact, the concept of ratifiability plays a key role in

Aumann's (1987) argument that common knowledge of Bayesian rationality implies a correlated equilibrium, although he does not isolate the concept and is not in contact with the philosophical literature on the subject.

Let us begin in the context of individual decision making. Here we can show that more than self-knowledge of Bayesian rationality requires that the decision-maker be sure that she will choose a rationalizable act. Suppose that at the time t_0, a decision-maker is undecided about which act to do, and has determined to do an experiment before deciding. The possible experimental results form a finite partition of her belief space, each of whose elements has positive probability at t_0. Then she is informed which element of the partition is the true one, and conditions on this information. Finally, the decision-maker decides by t_1 on an act which maximizes expected utility with respect to this posterior probability. We assume that the decision-maker is equipped with an appropriate Jeffrey-Savage space, with finite act and state partitions. Now let us also assume that the decision-maker *knows* at t_0 that she will choose an act which maximizes expected utility at t_1 (and knows which act she will choose if there is a tie). Then we can show that the decision-maker knows at t_0 that one of the acts ratifiable at t_0 will be chosen at t_1.

Ratifiability Lemma: If the decision-maker about to perform an experiment with a finite number of outcomes knows at t_0 that she will receive an experimental result, condition on it, and then choose an act which maximizes expected utility, and knows which act she will choose for every possible experimental result, then she knows that she will choose an act that is ratifiable at t_0.

Proof: Conditional on each member of the information partition e, there is an act, A such that $pr(A|e) = 1$ and A maximizes expected utility conditional on e. Conditioning on an act, A, is equivalent to conditioning on the union of the members, e, of the information partition such that $pr(A|e) = 1$. It is an algebraic property of Savage expected utility that if an act, A, maximizes Savage expected utility conditional on one member, e, of a partition and on another member, e', of that partition, then it maximizes Savage expected utility conditional on their union. If an act A has any prior probability of being chosen, it must maximize expected utility conditional on some member of the partition. Then by the foregoing algebraic property of Savage expected utility it must maximize expected utility conditional on itself; it must be ratifiable. By contraposition, non-ratifiable acts get probability 0 at t_0. Since the act partition is finite, the decision-maker is sure at t_0 that she will choose one of the ratifiable acts.

Notice that there is nothing in the proof of the ratifiability lemma that requires the "experiment" to be a laboratory experiment in the ordinary sense. It might just consist of sitting and watching, or—if deliberation is conceived of as generating new information—of just sitting and thinking. The theorem applies in the degenerate case of vacuous information—where the information partition has only one member which is the whole space. In this case, we learn nothing so our probability at t_1 is the same as our probability at t_0. Then at t_0 the decision-maker must already have an act which she is sure that she will do, and which maximizes expected utility at t_0.

Let us now consider a finite non-cooperative game played by Bayesian decision-makers. Here each player's state of the world consists of the combination of acts of all the other players. We can then conveniently assume a common Jeffrey-Savage space whose points consist of combinations of acts of all the players, over which all players have their relevant beliefs. Aumann is interested in a model which gives each player the kind of decision setup we have just considered. At time t_0 players are undecided about what to do. At time t_1, each has gotten private information as to the true member of some information partition, has conditioned on that information, and decided on a pure act which then has probability 1 for the actor in question. Aumann wishes to prove that here common knowledge of Bayesian rationality at t_0 implies that the players are at a correlated equilibrium at t_0. Common knowledge of Bayesian rationality is construed as implying that at t_0, each player is sure that she will maximize expected utility at t_1. It is also assumed that each agent knows how she will break ties. Thus, each agent has at t_0 for every element of her information partition, an act which has probability 1 conditional on that element and which maximizes expected utility conditional on that element.

Then, by the ratifiability lemma, each player is sure at t_0 that she will do a ratifiable act. By the definition of ratifiability, if each player were told privately at t_0 what ratifiable act she would do at t_1, she would have had no incentive to deviate. This means that the players are at a *subjectively correlated equilibrium*[13] at t_0. If, in addition, they share the same probability at t_0, they are at an objectively correlated equilibrium. If, furthermore, they already know what they are going to do, they are at a Nash equilibrium.

Ratifiability is an equilibrium concept for individual decision making. Together with various degrees of common knowledge in game theoretic situations, it generates the main equilibrium concepts of the theory of games.

6. THE STRENGTH OF SELF-KNOWLEDGE

The assumptions of self-knowledge required by the ratifiability lemma may appear to be quite modest but they have dramatic consequences. As our first illustration, let us see how a decision-maker with such self-knowledge would probabilize a Newcomb problem. The following is stipulated:

	Payoff		*Probabilities*	
	State 1	State 2	State 1	State 2
Act 1	1,000,000	0	x	0
Act 2	1,001,000	1,000	0	y

Now the sort of reasoning that Jeffrey had in mind when he introduced the concept of ratifiability goes through. By the ratifiability lemma, the decision-maker knows at t_0 that he will do a ratifiable act. But there is only one ratifiable act here. That is Act 2. So the decision-maker's initial probabilities at t_0 must already be $x = 0$ and $y = 1$. Self-knowledge of expected utility maximization in a curious way prohibits the existence of values for x and y such that Jeffrey expected value maximization is well defined and conflicts with Savage expected utility maximization. Under these conditions of self-knowledge, the problematic cases for Jeffrey *vs* Savage are ones like the Nice Demon rather than Newcomb's problem.

The power and scope of assumptions of self-knowledge should not be underestimated. Self-knowledge carries with it a kind of self-reference, and flirts with paradox. Consider the case of the Mean Demon. (As a game, take the decision-maker to be playing a zero-sum game with a mean demon. Mixed strategies are not available.[14])

The Mean Demon

	Payoff		*Probabilities*	
	State 1	State 2	State 1	State 2
Act 1	- 50	0	x	0
Act 2	0	-100	0	y

In Newcomb's problem, the assumptions of self-knowledge constrained the possible probabilities at t_0 so that only one initial probability was possible. Here these assumptions overconstrain the initial probabilities. There is no ratifiable act, so the assumptions of the ratifiability lemma cannot be met.

If we weaken the assumptions of self-knowledge slightly, by omitting the seemingly innocuous assumption that the decision-maker knows how she will break ties, then the remaining assumption of knowledge of expected utility maximization becomes barely consistent. Suppose that no matter what information comes in, at t_1 the values of x and y are $2/3$ and $1/3$ respectively. (Since coherence requires that prior probability be the expectation of posterior probability, the values of x and y at t_0 must be the same. We are in the degenerate case of an experiment, where no relevant information is produced.) Then at t_1 both acts maximize expected utility, and knowledge at t_0 that one will do something that maximizes expected utility at t_1 is just knowledge that one will do something. You can check that these are the *only* values for x and y at t_0 and t_1 which are consistent with the weakened assumption. In contrast, the bare assumption that the decision-maker will maximize expected utility (although she may not know it) is consistent with any probability values for x and y.

The question of the existence of ratifiable acts—and thus of the self-knowledge assumptions that guarantee existence—is a delicate one which calls for careful examination. Such an investigation would not be complete unless it paid attention to the possible deliberational processes which could take the decision-maker from a state of indecision to a state of decision. A myopic deliberator, who does not think about the moment of truth, may simply fail to converge to a decision in certain nasty cases.[15] A deliberator with foresight, who recognizes the problem as one which has no ratifiable decision, may as well forget about ratifiability and opt for the act with highest current expected utility,[16] notwithstanding the anticipated regrets.

7. CONCLUSION

The main significance of ratifiability does not lie in its use in handling causal pathology within the framework of *The Logic of Decision*. A proper implementation of ratifiability within that framework requires the identification of partitions which do the work of Savage's distinction between acts and states. Given those partitions, one can recover choiceworthiness most simply as maximum Savage expected utility. The real importance of Jeffrey's introduction of the concept of ratifiability into current philosophical discussion of individual rational decision is that it has opened up a rich array of topics for investigation: the status of equilibrium as a rationality concept, the nature of the deliberational process, the connections between individual decision theory and the theory of games, and the consequences of self-knowledge for rational decision-makers.

8. GUIDE TO THE LITERATURE

Jeffrey's system was introduced in Jeffrey (1965) as a more general framework than that of Savage (1954). Questions of interpretation arising from the inclusion of acts in the probability space in this system are discussed by Sneed (1966), Spohn (1977), Jeffrey (1977), and Shin (1989; forthcoming). The divergence between expected utility in the sense of Savage and expected value in the sense of Jeffrey is discussed in Gibbard and Harper (1981). The connection between maximization of Jeffrey expected value and the concept of Stackelberg equilibrium is pointed out in Walliser (1988). Gibbard and Harper (1981) advance an alternative "causal decision theory" formulated in terms of probabilities of subjunctive conditionals. For a survey of causal decision theories and a demonstration of their essential unity with each other and with Savage, see Skyrms (1980; 1984) and Lewis (1980). For a representation theorem for generalized causal decision theory, see Armendt (1986, 1988). For an explicit semantics for the relevant counterfactuals in normal form games, see Shin (1989).

In the second edition of *The Logic of Decision* (1983), Jeffrey gives an informal definition of ratifiability: "A ratifiable decision is a decision to perform an act of maximum estimated desirability relative to the probability matrix an agent thinks he would have if he finally decided to perform that act." The definition used in this essay is one way of realizing this basic idea. The definition of ratifiability proposed by Harper (1984; 1986; 1988) is equivalent to that given in section 2 of this essay, provided that the counterfactual in Harper's definition is given reasonable truth conditions in terms of the Savage states. This notion is called *stability* by Rabinowicz (1988). Walliser (1988) takes it as a natural interpretation of Jeffrey's informal definition. The same notion is independently introduced by Aumann (1987) to demonstrate the connection between *ex ante* optimality and his (1974) notion of correlated equilibrium. The significance of ratifiability is discussed by Eells (1988a; 1984b; 1985); Eells and Harper (forthcoming); Harper (1984; 1986; 1988); Rabinowicz (1988; forthcoming); Richter (1984; 1986); Shin (1989; forthcoming); Sobel (1986); and Weirich (1985; 1986).

For the relevant notions of dynamic coherence see Goldstein (1983), van Fraassen (1984), and Skyrms (1987a; 1987b; 1990). Some models of the deliberational process are to be found in Eells (1984b), Jeffrey (1988), and Skyrms (1988; 1990).

NOTES

1. This essay was delivered at a conference on probability and rational decision in honor of Richard Jeffrey at Dunwalke, New Jersey, in September 1989. Research was partially supported by the National Science Foundation.

2. Related ideas had been discussed by Ellery Eells (1982).

3. Spohn argues vigorously against this move.

4. As do a host of interrelated "causal decision theories." For simplicity and familiarity this essay will rely on Savage as a member of this group.

5. Jeffrey (1965, 19).

6. Ibid., 84. But also see the discussion of probabilistic acts on pp. 177–79 which I will not attempt to model here.

7. See the discussion in Skyrms (1985).

8. In particular, Jeffrey and Eells.

9. Since the probability of the act chosen, A, is here equal to 1, we have for each K_i, $\mathrm{pr}(K_i) = \mathrm{pr}(K_i|A)$, so:

   Jeffrey Expected Value $= \Sigma_i \mathrm{pr}\,(K_i|A)\ U(K_i \,\&\, A) = \Sigma_i \mathrm{pr}(K_i)\ U\,(K_i \,\&\, A) =$ Savage Expected Utility.

10. If we were to model deliberation where the probability of the act selected fell short of 1 at t_1, there would be ample room for a Newcomb-type of spurious correlation to remain at the close of deliberation. In the second edition of *The Logic of Decision*, Jeffrey reports van Fraassen's dramatization of this kind of possibility in the case of Prisoners' dilemma with a clone.

11. See Jeffrey (1988).

12. This may not be a decisive argument against the program of showing that for certain special kinds of rational decision-makers, Jeffrey and Savage coincide in the limit. Perhaps the program can argue that such decision-makers should not be allowed to have the kind of prior given in the example.

13. For the definition of this equilibrium concept, see Aumann (1974; 1987). Limitations of space preclude a detailed discussion of game theoretic equilibrium concepts here.

14. Of course, if we expand the acts to include costless implementation of arbitrary random strategies, and preclude any correlation between the Savage states and *these* mixed acts, we get a ratifiable act. But perhaps implementation of a random strategy might carry with it a cost large enough to make any pure strategy preferable. Perhaps the mean demon might be

able to predict whether the decision-maker will randomize. Perhaps the mean demon could predict the outcome of the randomization. Solution of the existence problem is not something that we are entitled to take for granted.

15. Once the problem of convergence of deliberation is raised, it may be relevant even in cases where a ratifiable act does exist. For a problematic individual decision problem, see the shell game in Skyrms (1984). Similar examples have been discussed by Rabinowicz (1986) where mixed strategies are unavailable:

	C1	C2	C3
R1	1,3	2,0	3,1
R2	0,2	2,2	0,2
R3	3,1	2,0	1,3

The only Nash equilibrium is a (R2,C2). Therefore, under appropriately strong conditions of common knowledge, R2 and C2 are the only ratifiable options for Row and Column. But R2 is weakly dominated by both R1 and R3 and C2 is weakly dominated by both C1 and C3. For the application of dynamic deliberation models to this game, see Skyrms (1990).

16. As advocated by Rabinowicz (1988). See also the discussion in Skyrms.

REFERENCES

Armendt, B. 1986. "A Foundation for Causal Decision Theory." *Topoi* 5: 3-19.

Armendt, B. 1988. "Conditional Preference and Causal Expected Utility." In *Causation in Decision, Belief Change, and Statistics*, edited by W. Harper and B. Skyrms, 3-24. Dordrecht.

Aumann, R. J. 1974. "Subjectivity and Correlation in Randomized Strategies." *Journal of Mathematical Economics* 1: 67-96.

Aumann, R. J. 1976. "Agreeing to Disagree." *The Annals of Statistics* 4: 1236-39.

Aumann, R. J. 1987. "Correlated Equilibrium as an Expression of Bayesian Rationality." *Econometrica* 55: 1-18.

Eells, E. 1982. *Rational Decision and Causality*. Cambridge.

Eells, E. 1984a. "Causal Decision Theory." *PSA 1984*, vol. 2, edited by P. Asquith and P. Kitcher, 177-200. East Lansing, Mich.

Eells, E. 1984b. "Metatickles and the Dynamics of Deliberation." *Theory and Decision* 17: 71-95.

Eells, E. 1985. "Weirich on Decision Instability." *Australasian Journal of Philosophy* 63: 473-78.

Eells, E., and W. Harper. Forthcoming. "Ratifiability, Game Theory and the Principle of Independence of Irrelevant Alternatives." *Australasian Journal of Philosophy.*

Gibbard, A., and W. Harper. 1981. "Counterfactuals and Two Kinds of Expressed Utility." In *Ifs*, edited by Harper et al., 153-90. Dordrecht.

Goldstein, M. 1983. "The Prevision of a Prevision." *Journal of the American Statistical Association* 78: 817-19.

Harper, W. 1984. "Ratifiability and Causal Decision Theory: Comments on Eells and Seidenfeld." In *PSA 1984*, vol. 2, edited by P. Asquith and P. Kitcher. East Lansing, Mich.

Harper, W. 1986. "Mixed Strategies and Ratifiability in Causal Decision Theory." *Erkentniss* 24: 25-36.

Harper, W. 1988. "Causal Decision Theory and Game Theory: A Classic Argument for Equilibrium Solutions, a Defense of Weak Equilibria, and a New Problem for the Normal Form Representation." In *Causation in Decision, Belief Change, and Statistics*, edited by W. Harper and B. Skyrms, 25-48. Dordrecht.

Jeffrey, R. 1965. *The Logic of Decision.* New York. 2d rev. ed. 1983. Chicago.

Jeffrey, R. 1977. "A Note on the Kinematics of Preference." *Erkentniss* 11: 135–41.

Jeffrey, R. 1981. "The Logic of Decision Defended." *Synthese* 48: 473–92.

Jeffrey, R. 1988. "How to Probabilize a Newcomb Problem." In *Probability and Causality,* edited by J. Fetzer, 241–51. Dordrecht.

Lewis, D. 1980. "Causal Decision Theory." *Australasian Journal of Philosophy* 59: 5–30.

Moulin, H. 1986. *Game Theory for the Social Sciences.* New York.

Rabinowicz, W. 1985. "Ratificationism without Ratification." *Theory and Decision* 19: 171–200.

Rabinowicz, W. 1988. "Ratifiability and Stability." In *Decision, Probability and Utility,* edited by R. Gardenfors and N. Sahlin, 406–25. Cambridge.

Rabinowicz, W. Forthcoming. "Stable and Retrievable Options." *Philosophy of Science.*

Richter, R. 1984. "Rationality Revisited." *Australasian Journal of Philosophy* 62: 392–403.

Richter, R. 1986. "Further Comments on Decision Instability." *Australasian Journal of Philosophy* 64: 345–49.

Savage, L. J. 1954. *The Foundations of Statistics.* New York.

Shin, H. S. 1989. "Counterfactuals and a theory of equilibrium in games." Read at the workshop on Knowledge, Belief and Strategic Rationality, Castiglioncello, Italy.

Shin, H. S. Forthcoming. "Two Notions of Ratifiability and Equilibrium in Games." In *Essays in the Foundations of Decision Theory*, edited by M. Bacharach and S. Hurley.

Skyrms, B. 1980. *Causal Necessity.* New Haven, Conn.

Skyrms, B. 1984. *Pragmatics and Empiricism.* New Haven, Conn.

Skyrms, B. 1985. "Ultimate and Proximate Consequences in Causal Decision Theory." *Philosophy of Science* 52: 608–11.

Skyrms, B. 1987a. "Dynamic Coherence." In *Foundations of Statistical Inference,* edited by I. B. MacNeill and G. J. Umphrey, 233–43. Dordrecht.

Skyrms, B. 1987b. "Dynamic Coherence and Probability Kinematics." *Philosophy of Science* 54: 1–20.

Skyrms, B. 1988. "Deliberational Dynamics and the Foundations of Bayesian Game Theory." In *Epistemology* [Philosophical Perspectives v. 2], edited by J. E. Tomberlin, 345–67. Northridge.

Skyrms, B. 1990. *The Dynamics of Rational Deliberation.* Cambridge, Mass.

Sneed, J. D. 1966. "Strategy and the Logic of Decisions." *Synthese* 16: 270–83.

Sobel, J. H. 1986. "Defenses against and conservative reactions to Newcomb-like problems: Metatickles and Ratificationism." In *PSA 1986,* vol. 1, edited by A. Fine and P. Machamer, 342–51. East Lansing, Mich.

Sobel, J. H. Forthcoming. "Maximization, Stability of Decision and Actions in Accordance with Reason." *Philosophy of Science.*

Spohn, W. 1977. "Where Luce and Krantz Do Really Generalize Savage's Decision Model." *Erkentniss* 11: 113–34.

van Fraassen, B. 1984. "Belief and the Will." *Journal of Philosophy* 81: 235–56.

Walliser, B. 1988. "A simplified taxonomy of 2 X 2 Games." *Theory and Decision* 25: 163–91.

Weirich, P. 1985. "Decision Instability." *Australasian Journal of Philosophy* 64: 465–72.

Weirich, P. 1986. "Decisions in Dynamic Settings." In *PSA 1986,* edited by A. Fine and P. Machamer, 438–49. East Lansing, Mich.

Weirich, P. 1988. "Hierarchical Maximization of Two Kinds of Expected Utility." *Philosophy of Science* 55: 560–82.