BRAD ARMENDT

CONDITIONAL PREFERENCE AND CAUSAL EXPECTED UTILITY

I am going to describe how we can make use of the idea of conditional preference to provide a foundation for Brian Skyrms' version of causal decision theory (*K*-expectation, or causal expected utility, decision theory). The foundation has the following two virtues: first, it puts the theory on an equal footing with competitors for which foundations have already been given; second, this foundation will provide the basis for a reply to the most serious objections made so far against causal decision theory, and against Skyrms' version of it in particular (more about this below).[1] I will only say a little about the other versions of causal decision theory developed by Gibbard and Harper (1976), Lewis (1981), and Sobel (1978). There are interesting differences between the theories, but in spirit they are very much alike. The foundation and the replies are immediate payoffs of using conditional preferences, but I think the treatment of conditional preference sketched below is also interesting for its own sake.

The foundation I will describe consists of (1) a set of axiomatic conditions on rational preference systems, and (2) the derivation of a representation theorem which shows that for any preference system satisfying the axioms there exist a probability measure *P* and utility functions *U* which represent the preferences, and which are related by the theory's general expected utility rule. The theorem has standard uniqueness results: for each preference system, *P* is uniquely determined and *U* is unique up to positive linear transformation. I should make it clear from the beginning that this foundation relies on formal results of Fishburn's (1973). My alterations of his formal theory are small. I do reinterpret his theory somewhat. The application to causal decision theory is new.

CAUSAL DECISION THEORY

Causal decision theories were proposed in reaction to the perceived failure of so-called *V*-maximization theories to adequately deal with problems like the Newcomb problem and the Twin Prisoners'

William L. Harper and Brian Skyrms (eds.), Causation in Decision, Belief Change, and Statistics, II, 3–24.

Dilemma[2]. In Richard Jeffrey's V-maximization theory the idea that actions are to be evaluated by their expected values takes the form

$$\text{(J)} \qquad V(A) = \sum_i P(C_i/A)\, V(A \,\&\, C_i),$$

for any A and any partition $\{C_1, \ldots, C_n\}$. This theory has the nice feature that no initial classification of acts, states, and consequences need be made: all appear as members of a Boolean algebra of propositions, and there is no requirement that acts and states of the world must be independent. So, though the notation suggests that the rule evaluates actions in terms of consequences, rule (J) is quite general, holding for any proposition A and any partition of C's. Iteration of (J) yields (J2) which might be useful to an agent who is better able to estimate the probabilities and values appearing in this rule than the ones appearing in (J):

$$\text{(J2)} \qquad V(A) = \sum_j P(B_j/A) \sum_i P(C_i/A \,\&\, B_j)\, V(A \,\&\, C_i \,\&\, B_j),$$

for any A and any partitions $\{B_1, \ldots, B_m\}$, $\{C_1, \ldots, C_n\}$. V-maximization is usually a good theory: most of the time an agent using V-maximization will accurately evaluate his alternatives. And I will show later that the utility rule given by the foundation I will present agrees with V-maximization in the many cases V-maximization gets right.

But (J2) can lead an agent astray in the problems we might call *causal counterexamples* to V-maximization. The best-known of these are the Newcomb problem and the Twin Prisoners' Dilemma, but here I will illustrate the problem for V-maximization with Fisher's smoking gene hypothesis.

Fisher's Smoking Gene Example

I believe that my disposition to smoke cigarettes and my disposition to contract cancer are genetically influenced by the same factor, which accounts for the statistical correlation between smoking and cancer. I believe that I cannot influence my genetic makeup, and that smoking itself is not harmful. I would enjoy smoking and attach small positive utility to the pleasure I would derive from smoking this cigarette. I attach large negative utility to contracting cancer. If I believe that the causal connections and statistical correlations are strong enough, and if I disvalue contracting cancer enough, compared to how much I value the pleasure of smoking, then (J) and V-maximization will lead me to the decision to refrain from smoking, in order to minimize the chances I have the gene and contract cancer.[3]

But this answer is wrong. If I believe (1) that nothing I do now can contribute to my having or not having the genetic factor, and (2) that the genetic factor is the main cause of my contracting cancer while smoking itself is *not* a cause of cancer, then the correlation between smoking and having the gene is *not* a good reason for avoiding the pleasure I would get from smoking. To act otherwise is, as Gibbard and Harper (1976) remark, to knowingly act so as to produce *evidence* for a strongly desired state of affairs (absence of the gene), without in any way *producing* the desired state, *even when such action has significant cost*.

An agent who follows V-maximization and the reasoning sketched above wrongly evaluates his alternatives because he fails to use his most specific and relevant information (or beliefs) in the given situation, namely his full information about the causal structure of his decision problem. This is so because (the agent believes that) his action in no way causes any of the states B_j to obtain, although the action is correlated with the states. In these situations, doing A_s (smoking the cigarette) is not a cause but is a symptom of B_g's obtaining (having the gene), so the conditional degree of belief $P(B_g/A_s)$ is greater than the belief $P(B_g)$. V-maximization leads the agent to act so as to raise (at significant cost) the *epistemic* probability of the desirable state (not having the gene) even though the agent believes such action cannot cause the state to occur. V-maximization recommends this because it ignores the information about the causal independence of the action and the state (no dependence from action to state), while attending to the information about the epistemic dependence of one on the other.

The various causal decision theories are designed to correct V-maximization by incorporating the agent's beliefs about the absence of a causal connection from his action to the states in his evaluation of his action. The basic idea shared by all the causal decision theories is that if the agent believes that the world may be in one of several states (or have one of several structures) whose occurrence are not causally influenced by his action, and which each affect the chances the consequences have of being caused by the action (or which each affect the values of the consequences), then he should evaluate his action this way: For each of the possible states or structures, find the value the action has if that state holds; then find the *overall* value of the action by taking a weighted average of these values, using as weights the degrees of belief in each state's being the actual one. The theories differ in their description of the states and structures to be considered and in their analysis of the action's value for each state. The theories of Gibbard and Harper (1976) and Lewis (1981) both recommend that the agent consider counterfactual conditionals describing possible causal patterns the world might have that are relevant to the actions and consequences in question.[4] Which of these causal patterns obtains is taken to be outside the agent's control, and in both theories the agent is told to weight the values he gives to the possible consequences by his degrees of belief in the competing causal counterfactual conditionals. The appropriate conditionals are *causal* in the sense that in the smoking gene problem, for example, the agent would be expected to assign a substantial degree of belief to "I smoke $\Box\!\!\rightarrow$ I enjoy my cigarette," but a very low degree of belief to "I smoke $\Box\!\!\rightarrow$ I get cancer," given that he believes that his smoking does not cause him to have the gene, and that the only causal connection between the smoking and getting cancer is through this genetic factor. The appropriate causal counterfactuals are not "backtracking counterfactuals."

That is all I will say here about the other versions of causal decision theory. Skyrms' version is not formulated in terms of counterfactual conditionals, and it is to his theory that I will now turn.

Consider the smoking gene example and recall that the V-maximization formula (J2) does not incorporate a correct summary of my beliefs about the situation. The $P(B_j/A)$'s which appear in (J2) will reflect my belief that smoking is correlated with having the gene, but not my belief that smoking does not cause me to have it. We might delete the misleading influence that the correlation contributes to the evaluation of act A by replacing these conditional probabilities with the simple unconditional degrees of belief $P(B_j)$'s. If we take note of this alteration of our expected utility rule by writing the utility function U, and we note the special requirement on the states B_j by writing them as K_j, (J2) then becomes

$$\text{(S)} \qquad U(A) = \sum_j P(K_j) \sum_i P(C_i/A \;\&\; K_j)\, U(A \;\&\; C_i \;\&\; K_j).$$

(S) will agree with (J2) and V-maximization whenever the states K

which influence the chances or utilities of the consequences are statistically independent of the agent's choice of action. But when the states and the choice are not statistically independent, (J2) and (S) will very likely give different values to A, and these differences may lead to different recommendations when the agent chooses the alternative with the highest utility. It is important to note that our justification for suggesting (S) is that the states K_j are believed *causally relevant* to the action A's production of its possible consequences C_i and they are believed *causally independent* of the action (no dependence from action to state). If the K_j's satisfy those conditions, but not otherwise, then we correctly summarize the agent's beliefs by using (S). Skyrms calls partitions which in a given decision situation describe the causally relevant (possible) states of the world which are outside the agent's influence *K-partitions*. Skyrms' version of causal decision theory recommends that the agent choose the act which has maximum causal expected utility, or K-expected utility, given by (S).[5]

OBJECTIONS TO CAUSAL EXPECTED UTILITY THEORY

There has been considerable debate about whether or not we need to adopt causal decision theory in order to make the best choices in problems like these. Sophisticated V-maximization theories have been developed, and their defenders have argued that causal decision theory is superfluous. This view is mistaken, but I will not discuss it here.[6] Other criticisms have been made of Skyrms' causal expected utility theory that I *shall* consider.

It is easy to see that the key to applying the theory in the situations where it is most needed is the identification of an appropriate K-partition. An agent needs to find a set of propositions $\{K_j\}$ which are such that (a) their use in (S) is correct for the problem confronting him; that is, when they are used in (S) his evaluation of the action incorporates all his information/beliefs relevant to the problem; and (b) they are practical – the agent needs to have some idea what the values of the degrees of belief and utilities appearing in (S) actually are. As I said above, Skyrms (1979) has described the K-partition appropriate to a given problem as a partition of maximally specific descriptions of the factors outside the agent's influence which are causally relevant to the outcomes (that he cares about) of the alternatives available to him.

Now the theory has been criticized on the grounds that the appropriate K-partitions for decision problems must be carefully selected (which is true), that general use of the theory seems to depend upon a general way of finding an appropriate K-partition, and that this task requires an understanding on the agent's part of the relations "state S is outside my influence" and "state S is causally relevant to act A's having outcome C" that is more subtle than is reasonable to require for a useful decision theory, even a normative one. Further, it has been said that if Skyrms' theory lacks a representation theorem, it lacks a theoretical guarantee that it is broadly applicable – a guarantee possessed by other decision theories, including V-maximization.[7] The axiomatic system and representation theorem I shall describe provide a direct response to the latter complaint about the theory. We shall also find an answer to the former objections concerning the problem of the selection of K-partitions.

It is worth pointing out that for an agent with a given decision problem the task of selecting an appropriate K-partition is an empirical one. And a K-partition is correct for a given decision problem by being correct for the agent's *beliefs* and *preferences* about the problem. The causal decision theories are more successful than V-maximization by better incorporating the agent's beliefs, particularly his causal beliefs, about his problem into his act evaluations – the Newcomb game and the smoking gene example would be no problems at all for V-maximization if the agent did not have the beliefs ascribed to him in those situations. This is not a point on which the critics of causal expected utility theory go wrong, but it is worth mentioning because it leads to the idea that we should look to the agent's preference system, which underlies his beliefs and desires, if we want to describe the selection of appropriate K-partitions for his decision problem. In what follows, I will be able to state sufficient conditions for the propriety of K-partitions in terms of their behavior in the agent's preference system. And it will turn out that these conditions correspond quite well with Skyrms' description given above.

PREFERENCE AND CONDITIONAL PREFERENCE

The descriptions of the entities for which our agent is supposed to have preferences[8] are given, first, by propositions which may describe acts, states, or consequences, and second, by mixtures of propositions. So the agent may have in his system preferences for going swimming, for its raining, for catching a cold, and so on. These preferences are represented in our system by corresponding propositions "I go swimming," "It will rain tomorrow," "I catch a cold," and so on. In addition, mixtures or gambles on propositions will be taken to be descriptions of objects of preference. So also appearing in our system will be mixtures like $0.6(R) + 0.4(-R)$, where R is "It will rain tomorrow;" and $0.2(S \& C) + 0.5(S \& -C) + 0.3(-S)$, where S is "I go swimming," and C is "I catch a cold." Such mixtures can be thought of as lotteries, where the mixing coefficients give the odds on each of the possible outcomes. This system will assume that the agent has a rich set of these preferences for mixtures (including mixtures of mixtures, and so on). They form a *mixture set*, almost in the sense of Herstein and Milnor (1953): the exception is that here mixing is restricted to sets of propositions which form partitions, rather than being closed for all sets of propositions whatsoever. (This differs from Herstein & Milnor, who take the basic elements of mixture sets to be incompatible payoffs; there is no need in their system to further specify that mixtures be on incompatible outcomes). This modification of Herstein–Milnor's (and Fishburn's) theory will be discussed further below. Two points that I want to bring out here are that (1) Fishburn's theory and my adaptation of it follow, as do many other decision theories, the von Neumann and Morgenstern approach to utility in that the preference systems are taken to include mixtures or lotteries;[9] (2) this theory deviates from Fishburn's and most other theories, however, in its use of *propositions* and mixtures of *propositions*, rather than the more common practice of interpreting preferences as preferences for nonpropositional acts, either primitive or regarded, for example, as functions from states to consequences; in this respect it follows the Jeffrey–Bolker "monoset theory".

The set of preferences so far described, and indeed the full set of preferences described below, are assumed to be ordered by a relation $\geqslant$, interpreted as "is at least as preferable as." The entire system of ordered (rational) preferences will be assumed to satisfy a set of axiomatic conditions that I will present below; notice that the ordering assumption includes requirements that preference is transitive, and that any pair of elements of the ordering are comparable.

Now the system described so far is further enriched by the addition of *conditional preferences*. The agent may entertain an hypothesis about the world, and under that hypothesis he may find that his preference-attitudes toward various states, acts, etc. differ from the preference-attitudes he directs toward those states, acts, etc. under other hypotheses, or under the trivial hypothesis T (which, for convenience, is how I will regard unconditional preferences). That is, under one hypothesis H the agent may rank his preferences (conditional on H) quite differently from the ranking of his unconditional preferences, or the ranking of his preferences conditional on hypotheses J, K, ... Of course, conditional hypotheses need not *always* alter the ranking of preferences. For example, under my hypothesis that it rains here this afternoon, my preference for swimming this afternoon at the local beach is ranked below my unconditional preference for swimming this afternoon at that beach. But under the hypothesis that it snows today in Tibet, my preference for swimming this afternoon is not perturbed.

I shall write P,Q and m,Q for the conditional preferences for P and for m, under the hypothesis that Q. As I noted just above, I shall regard unconditional preferences as preferences conditional on the tautology T, and sometimes will write them P,T. A conditional preference, I emphasize, is not a preference for a conditional proposition. And P,Q is to be understood as the agent's *present* preference for P, under his hypothesis that Q – not (necessarily) as the preference he *would have* for P if Q were true, or if he were to *believe* Q true.

Sometimes a conditional preference may have the same ranking as an unconditional preference for a corresponding conjunction. The swimming examples above may well be illustrations of this; my preference for swimming, under my hypothesis that it rains, may be ranked the same as my unconditional preference for the conjunction of swimming and its raining. And similarly for my preference under my hypothesis about the snow in Tibet. But this is not always the case. Consider the preferences M,H and H,M where M is "I have medical insurance," and H is "I am hospitalized for a serious illness." My preference for having medical insurance, under the hypothesis that I am hospitalized, is considerably greater than my preference for being hospitalized, under the hypothesis that I have insurance. Of course, if both preferences were ranked with the corresponding conjunction, then I would be indifferent between them. So we will not assume that conditional preferences are always ranked with corresponding conjunctions, though in many cases it may turn out that they are.

What about preferences of the forms P,P and $-P,P$? I regard such conditional preferences as well-formed and non-trivial; they play a significant role in the description of the proper use of the causal expected utility rule (S). Under his hypothesis that the actual world is a P-world, the agent has preferences for its being a P-world and for its being a $(-P)$-world. It is important to keep in mind when considering $-P,P$ that this is not a preference for $-P$ & P, and it is certainly not a belief in $-P$ & P. One may have a well-defined unconditional preference for a proposition one strongly believes false, and one may have a conditional preference for a proposition assumed false – not just because the proposition may not be assumed to be known false, but mainly because preference and desire may be directed toward propositions in which one does not believe. I may desire that Pegasus be alive and willing to carry me wherever I please; and I may still desire this, under my hypothesis that Pegasus never lived in the actual world.

It will often turn out that making the hypothesis P (or the hypothesis $-P$) does *not* perturb an agent's preference for P. Let W be "Candidate C wins the election," and let S be "I go swimming this afternoon;" it is quite plausible that I be indifferent between W, W,W, and $W,-W$; and also between S, S,S, and $S,-S$. But consider the Newcomb problem, letting A_2 be "I take both boxes," and the smoking gene problem, letting A_s be "I smoke the cigarette." In the former case, it is highly plausible that A_2,A_2 is ranked below A_2, and $A_2,-A_2$ is ranked above A_2. Under the hypothesis that I take both boxes, my preference for taking both boxes is diminished, since worlds in which I take both boxes are worlds where an empty opaque box is likely. And since in worlds in which I do not take both boxes a filled opaque box is likely, the hypothesis that I do not take both raises my preference for taking them. For similar reasons, in the smoking gene example I prefer $A_s,-A_s$ to A_s to A_s,A_s. Another example that reinforces the idea that preferences like A,A and $A,-A$ make nontrivial sense is the story of the man who met death in Damascus, from Gibbard and Harper (1976): when the story is elaborated in the way Gibbard and Harper do, it is clear that the man prefers $A,-A$ to A to A,A, where A is "I go to Aleppo." In the theory that follows we will see that conditional preferences like A,A and $A,-A$ are important. The failure of an act proposition A to be ranked equally to A,A is an indication that use of the causal expected utility rule (S), rather than a simpler special case of it, is appropriate in evaluating A.

Two further points about conditional preference: First, though the examples so far mentioned have all been of the form P,Q where P is a proposition, conditional preferences for mixtures of the form m,Q are also included in the agent's preference system. But *hypotheses* are always taken to be propositions rather than mixtures; under a particular hypothesis H the agent's preferences form a mixture set (in our sense, mixtures only on partitions) written $\mathbf{M_H}$. Second, notice that mixtures are only formed under single conditional hypotheses; the agent is *not* assumed to have preferences like $0.6(P,Q) + 0.4(-P,R)$. Luce and Krantz (1971) have a well-known theory in which disjunctive preferences, each disjunct conditional on different hypotheses, are assumed to be in the preference system. And Balch and Fishburn (1974) present such a theory with mixtures like $0.6(P,Q) + 0.4(-P,R)$. I regard such preferences as intuitively very problematic and formally unnecessary.

The remarks made so far clearly amount to less than a complete account of conditional preference, but at this point I shall proceed with the description of the theory's preference systems and representation theorem. I claim that the assumptions made by this theory about conditional preference are actually not very strong, and that whatever correct account of conditional preference emerges from the intuitions I have appealed to will be consistent with those assumptions.[10]

DERIVATION OF CAUSAL EXPECTED UTILITY THEORY

I will discuss the axioms for rational preference systems and the representation theorem soon. Let us look ahead in this section and see how the results stated there provide a foundation for causal expected utility decision theory. Fishburn's 1973 representation theorem shows that a preference system $\langle \mathbf{X}, \succcurlyeq \rangle$ which satisfies the axioms to be stated below can be represented by a probability measure P and utility function U. U is linear and order-preserving, P is unique, U is unique up to positive linear transformation, and the following utility rule holds for all x in $\mathbf{X}$:

$$\text{(F1)} \qquad U(x,B\vee C) = P(B/B\vee C)\,U(x,B) + P(C/B\vee C)\,U(x,C)$$

whenever x,B and x,C are in $\mathbf{X}$, and B and C are incompatible.[11] And for A and a partition of propositions $B_1, B_2, \ldots, B_n$ such that A,B_i is in **X** for each i, iterated application of (F1) gives us the rule

$$\text{(F2)} \qquad U(A) = \sum_i P(B_i)\, U(A,B_i).$$

If U is a utility function and P the probability measure which represent my preferences, (F2) says that, for example, the utility I attach to swimming this afternoon $U(S)$ is equal to my degree of belief in rain $P(R)$ times the utility of swimming under the hypothesis that it rains, plus $P(-R)$ times the utility of S under the hypothesis $-R$. In the smoking gene problem, the utility I attach to smoking, A_s, is equal to my degree of belief that I have the gene, $P(G)$, times the utility of A_s under the hypothesis G, plus $P(-G)$ times $U(A_s,-G)$. This agrees with causal decision theory's evaluation of A_s, since the weights are *not* $P(G/A_s)$ and $P(-G/A_s)$ as in V-maximization theory.

When we compare (F2) with the V-maximization rule (J) it is clear that they agree when for each i, $P(B_i) = P(B_i/A)$ and $U(A,B_i) = U(A \,\&\, B_i)$. Since the function U is order-preserving, the latter conditions hold when I am indifferent between A,B_i and $A \,\&\, B_i$. So if I believe the states B_i are statistically independent of A, and my conditional preferences for A under the hypotheses B_i are the same as my unconditional preferences for the conjunctions $A \,\&\, B_i$, then evaluation of A using V-maximization will agree with this theory's evaluation of A.

If A is an act proposition, however, we are likely to want to evaluate it in terms of its possible *consequences*, rather than to use some other partition of states. And in any interesting decision problem, of course, the statistical independence will *not* hold – I will believe that A's possible consequences *are* influenced by whether or not I do A. There is another direction we may take, though, in seeking cases where the two rules agree: If I am indifferent between A and A,A (doing A, under the hypothesis I do A), and if for all i, I am indifferent between $A \,\&\, C_i$ and $A,(A \,\&\, C_i)$, then it is easy to see that (F2) agrees with rule (J). By order-preservation of U,

$$U(A) = U(A,A) = \sum_i P(A \,\&\, C_i/A)\, U(A,A \,\&\, C_i).$$

So by substitution,

$$= \sum_i P(C_i/A)\, U(A \,\&\, C_i).$$

So in such cases I may use the V-maximization rule (J) and I will get the correct value for $U(A)$. I think it turns out that these conditions, that $A \sim A,A$ and that A & $C_i \sim A,(A$ & $C_i)$, are satisfied in the ordinary decision situations where we expect V-maximization to give the correct evaluations of the available acts. The first says that my hypothesis that I do A does not perturb my preference for A, and it was discussed above. The second says that my preference for A, under my hypothesis that I do A and consequence C_i results, is equal to my unconditional preference for the conjunction of A and result C_i. In ordinary situations I think this simply requires that the C_i's be accurate descriptions of the possible consequences of doing A that I care about. Call it a requirement that the C_i's are *well-selected* consequence descriptions.

When does (F2) take the form of rule (S), the causal expected utility rule? Well, it does if it agrees with V-maximization, and if V-maximization in turn agrees with causal expected utility theory (i.e. with (S) for some appropriate K-partition). So in cases where that is true and where $A \sim A,A$, (F2) yields the causal expected utility rule which agrees (as it should) with V-maximization.

What about cases in which $A \nsim A,A$? Such cases include, as I have claimed above, the causal counterexamples to V-maximization. In these cases A is correctly evaluated by the causal expected utility rule when appropriate K-partitions can be found. How does (F2) yield the rule (S)? If we suppose that the agent is *not* indifferent between A and A,A, then for most state descriptions B_j we have little reason to expect that he will be indifferent between A,B_j and $A,(A$ & $B_j)$. But *some* state descriptions B_j may align those preferences after all: among the descriptions which do this are descriptions which make up appropriate K-partitions. For example, in the smoking gene story my hypothesis that I smoke the cigarette A_s devalues my preference for A_s. But consider my preferences A_s,G and $A_s,(A_s$ & $G)$ where G is "I have the gene." My preference for A_s, under my hypothesis that I have the gene, is *not* devalued by the additional hypothesis that I smoke. My preference for A_s was already devalued by the hypothesis G and (given the story assumed in this example) is not further affected by supposing that I do light up. Similarly for $-G$. Having or not having the gene is what matters in this decision problem, and once I suppose that I do or that I do not, my preference for A_s is fixed, with respect to the additional hypothesis A_s. This is true in general of propositions which are members of adequate K-partitions – the hypothesis that one of them holds in the actual world fixes the value of the act to the extent that whether or not the additional hypothesis that the act is done is added, the agent's preference for it remains the same.

So if we return to (F2), assume that $A,B_j \sim A,(A \,\&\, B_j)$, and then, in recognition of the special character of this partition of state descriptions, rewrite them as K_j's, we have

$$U(A) = \sum_j P(K_j)\, U(A,A \,\&\, K_j).$$

This will generate the rule (S) if we can find a partition of consequence descriptions which for each K_j satisfies our condition stated earlier for well-selected C_i's: If there is a partition $C_1, \ldots, C_n$ such that $A,(A \,\&\, C_i \,\&\, K_j) \sim A \,\&\, C_i \,\&\, K_j$ for all i,j, then we can use (F2) to analyze $U(A,A \,\&\, K_j)$ in terms of the possible consequences:

$$\begin{aligned} U(A,A \,\&\, K_j) &= \sum_i P(C_i/A \,\&\, K_j)\, U(A,A \,\&\, C_i \,\&\, K_j) \\ &= \sum_i P(C_i/A \,\&\, K_j)\, U(A \,\&\, C_i \,\&\, K_j). \end{aligned}$$

And so by substitution into the first equation of this paragraph:

$$U(A) = \sum_j P(K_j) \sum_i P(C_i/A \,\&\, K_j)\, U(A \,\&\, C_i \,\&\, K_j).$$

This is the causal expected utility rule (S).

So when $A \nsim A,A$ our theory endorses rule (S) for partitions $K_1, \ldots, K_m$ and $C_1, \ldots, C_n$ which satisfy

$$\begin{aligned} A,K_j &\sim A,(A \,\&\, K_j) \quad \text{and} \\ A,(A \,\&\, C_i \,\&\, K_j) &\sim A \,\&\, C_i \,\&\, K_j, \end{aligned}$$

for all i and j. This is a nice result since these are exactly the conditions that we should expect appropriate K and C partitions to satisfy. Skyrms describes the appropriate K-propositions as maximally specific descriptions of the factors the agent believes are outside his influence and causally relevant to those outcomes of his available alternatives which matter to him. When such a description K_j is made a conditional hypothesis, it may of course perturb the agent's preference for one of his alternatives A: $A_s,G > A_s$ in the smoking gene example. But once that hypothesis is made, the uncertain states of the world which may influence the outcome of the action (including the states which are correlated with but not caused by the action) are fixed, and the additional hypothesis that A is done should not further perturb the preference for A. The condition $A,K_j \sim A,(A \& K_j)$ captures the idea behind Skyrms' appropriate K-partitions, and it does so in the desirable way mentioned earlier: adequate K-partitions are picked out by reference to the way they behave in the agent's preference system.

Now any foundation for rational decision theory, even one which provides the nice results shown above, depends much for its adequacy on the assumptions about preference contained in its axioms. In the next section, I shall present the axioms for rational preference and the theorem. Discussion of the adequacy of the axioms can be found in Armendt (1983) and (1986a). I shall conclude with some sketchy remarks about the possibility of providing a different foundation – one more like the Jeffrey–Bolker foundation for V-maximization.

THE AXIOMS AND THEOREM

The structure which is interpreted as the agent's preference ranking is a collection of mixture sets each of whose basic elements are members of a Boolean algebra of propositions. The mixture sets are like Herstein–Milnor mixture sets, except that only mixtures of partitions are defined. Each mixture set corresponds to the agent's preferences under the hypothesis that some one element of $\mathscr{E}$ is true. We start with our set of propositions $\mathscr{E}$ that describe states, acts, and consequences. After deleting the contradictory proposition, we construct a set **M** of all mixtures or gambles on partitions of propositions in $\mathscr{E}'$ ($\mathscr{E}' = \mathscr{E} - \phi$). **M** is a large set containing all such mixtures, and $\mathbf{M} \times T$ is the set of all the agent's unconditional preferences.

A1. $\mathscr{E}$ is a Boolean algebra of propositions. And $\mathscr{E}' = \mathscr{E} - \phi$.

A2. **M** is a mixture set formed from $\mathscr{E}'$, i.e. $\mathscr{E}' \subset \mathbf{M}$, and:

a. *Closure under mixing*: For all $A,B \in \mathscr{E}'$ such that $A \& B = \phi$, and all $\alpha \in [0,1]$,

$$\alpha A + (1 - \alpha)B \in M;$$

and for all mixtures m on $A_1, A_2, \ldots, A_n$ and for all B disjoint from each A_i:

$$\alpha m + (1 - \alpha)A_i \text{ and } \alpha m + (1 - \alpha)B \text{ are in } \mathbf{M};$$

b. *Identity*: for all m_1, m_2 in $\mathbf{M}$, and for all $\alpha,\beta \in [0,1]$:

$$1(m_1) + 0(m_2) = m_1$$
$$\alpha m_1 + (1 - \alpha)m_2 = (1 - \alpha)m_2 + \alpha m_1$$
$$\beta[\alpha m_1 + (1 - \alpha)m_2] + (1 - \beta)m_2 = \alpha\beta m_1 + (1 - \alpha\beta)m_2.$$

The set $\mathbf{X}$ is the set which contains *all* the agent's preferences, conditional and unconditional. It contains all the elements of $\mathbf{M} \times T$ together with the elements of $\mathbf{M_P}$ for every proposition P in $\mathscr{E}'$. For each P in $\mathscr{E}'$, $\mathbf{M_P}$ is interpreted as the set of the agent's preferences for the elements of $\mathbf{M}$ under the hypothesis P. It is a (possibly improper) subset of $\mathbf{M} \times P$. It is assumed to be a mixture set in the sense given above. $\mathbf{X}$ is the union of all the $\mathbf{M_P}$'s and is ordered by the relation $\geqslant$. Notice that mixtures of conditional preferences are only allowed when they are formed under the same hypothesis: $\mathbf{X}$ contains $[\alpha P + (1 - \alpha)Q],R$ but not $\alpha P,R + (1 - \alpha)Q,S$.

A3. For all $P \in \mathscr{E}'$, $\mathbf{M_P} \subseteq \mathbf{M} \times P$ is a mixture set (see A2). Also, $\mathbf{M_T} = \mathbf{M} \times T$.

A4. The set $\mathbf{X}$ is the union of all the $\mathbf{M_P}$'s and is ordered by $\geqslant$.

We now come to a pair of axioms which are Herstein–Milnor's axioms for mixture set functions generalized to $\mathbf{X}$ (which is a union of mixture sets), and incorporating our restriction that mixtures be on disjoint propositions

A5. For all A, $B \in \mathscr{E}'$, all x,A, $y,A \in \mathbf{M_A}$ such that mixtures on x,A and y,A are defined, and all $z,B \in \mathbf{M_B}$, these sets are closed in $[0,1]$:

$$\{\alpha\colon (\alpha x + (1 - \alpha)y),A \geqslant z,B\} \quad \text{and}$$
$$\{\alpha\colon z,B \geqslant (\alpha x + (1 - \alpha)y),A\}.$$

A6. For all A, $B \in \mathscr{E}'$, all x,A, $y,A \in \mathbf{M_A}$, and for all z,B, $w,B \in \mathbf{M_B}$, if $x,A \sim z,B$ and $y,A \sim w,B$, and if the mixtures below are defined, then

$$[\tfrac{1}{2}x + \tfrac{1}{2}y],A \sim [\tfrac{1}{2}z + \tfrac{1}{2}w],B.$$

Setting $A = B$, axioms A5 and A6 are almost the Herstein–Milnor axioms, and in the absence of any restrictions on mixing, would imply that Herstein–Milnor's axioms are satisfied by the elements of mixture set $\mathbf{M_A}$ for every A. So then (via the Herstein–Milnor theorem) for every $\mathbf{M_A}$ there exists a real-valued utility function U_A which is linear and order-preserving over $\mathbf{M_A}$, and is unique up to positive linear transformation. The next axiom, A7, yields these results even in the presence of the restriction that only mixtures on partitions are defined:

A7. For all $A \in \mathscr{E}'$, for all $x,A \geqslant y,A$, there exist disjoint P and Q such that P,A, $Q,A \in \mathbf{M_A}$ and

$$P,A \geqslant x,A \geqslant y,A \geqslant Q,A.$$

Axiom 8 guarantees that whenever $A \,\&\, B = \phi$ and x,A and x,B exist, then so does $x,(A \vee B)$. Axiom 9 then states Fishburn's averaging principle. Axiom 10 is a non-triviality condition.

A8. For all A, $B \in \mathscr{E}'$ such that $A \,\&\, B = \phi$, and for all $x \in \mathbf{M}$, if $x,A \in \mathbf{M_A}$ and $x,B \in \mathbf{M_B}$, then $x,(A \vee B) \in \mathbf{M_{A \vee B}}$.

A9. For all x,A and x,B in X, if $A \,\&\, B = \phi$ and $x,A \geqslant x,B$, then $x,A \geqslant x,(A \vee B) \geqslant x,B$.

A10. For some x, $y \in \mathbf{X}$, $x > y$.

Axiom 11 is required to generate the comprehensive utility function U on $\mathbf{X}$ from the many U_P's on the $\mathbf{M_P}$'s. From the x and y whose existence it asserts, a gamble z on x and y (or on stand-ins for which the gamble is defined) will be found such that $z,A \sim z,B$. This axiom denies that there are two incompatible hypotheses A and B such that every element of $\mathbf{M_A}$ is preferred to the corresponding element of $\mathbf{M_B}$, and it is a fairly strong structural condition:

A11. For all A, $B \in \mathscr{E}'$ such that $A \,\&\, B = \phi$, there exist x, $y \in \mathbf{M}$ such that $x,A > x,B$ and $y,B > y,A$.

Axioms A1–A11 imply the existence of a utility function U on $\mathbf{X}$ which is linear, order-preserving, and unique up to positive linear transformation, and also the existence for each incompatible A and B in $\mathscr{E}'$ of unique non-negative real numbers which sum to 1, $P_{A \vee B}(A)$ and $P_{A \vee B}(B)$, such that

$$U(x,A \vee B) = P_{A \vee B}(A)\, U(x,A) + P_{A \vee B}(B)\, U(x,B),$$

for all x,A, $x,B \in \mathbf{X}$ (Fishburn, Th. 3). Axiom 12 is required to guarantee the additivity of the probabilities P_A and the chain condition: $P_C(A) = P_C(B)\, P_B(A)$, whenever $A \Rightarrow B \Rightarrow C$:

A12. For all $A,B,C \in \mathscr{E}'$ that are pairwise incompatible, if there is an $x \in \mathbf{M}$ such that $x,A \sim x,B$, then there is a $y \in \mathbf{M}$ such that exactly two of y,A, y,B, and y,C are indifferent.

Axioms A1–A12 imply the representation theorem:

THEOREM (Fishburn, 1973). *If $\mathscr{E}$, $\mathbf{X}$, and $\geqslant$ satisfy axioms* 1–12 *above, then there is a real-valued function U on* $\mathbf{X}$ *and a finitely-additive probability measure P_A on $\{A \,\&\, B : B \in \mathscr{E}\}$ for each $A \in \mathscr{E}'$ such that:*

i. *$x,A > y,B$ iff $U(x,A) > U(y,B)$, for all x,A and y,B in* $\mathbf{X}$;
ii. *$U(\cdot,A)$ is linear (as a function on $\mathbf{M_A}$) for each A in $\mathscr{E}'$;*
iii. *$P_C(A) = P_C(B)\,P_B(A)$ whenever $A \Rightarrow B \Rightarrow C$, $A \in \mathscr{E}$, and B, $C \in \mathscr{E}'$;*
iv. *$U(x,A \vee B) = P_{A \vee B}(A)\,U(x,A) + P_{A \vee B}(B)\,U(x,B)$ whenever x,A, $x,B \in \mathbf{X}$ and $A \,\&\, B = \phi$;*

furthermore, the P_A's are unique and U is unique up to positive linear transformation.

Clause (iii) of this theorem provides measures P_B even when $P_T(B) = 0$. Note that whenever $A \Rightarrow B$ and $P_T(B) > 0$, $P_B(A) = P_T(A/B)$. Clause (iv) of the theorem is the general form of our theory's rule (F1) relating the utilities of a proposition or mixture under different conditional hypotheses.

Proof: Fishburn's system does not include the restrictions that appear in A2 (a), the closure condition on mixing, which restrict mixing to *incompatible* propositions. (This restriction permits us to regard mixtures as always on partitions, since we can think of $\alpha P + (1 - \alpha)Q$ as $\alpha P + (1 - \alpha)Q + 0[-P \,\&\, -Q]$. Fishburn does not include such a restriction because he takes acts, rather than propositions, to be the object of preference.) Restricting the closure condition on mixing in this way leads to the introduction of Axiom 7, which is not Fishburn's, into our system. The proof of the theorem for our system is as in Fishburn (1973), with the following modifications:

First, one can check the proof of the Herstein–Milnor theorem to see that it applies to the sort of mixture sets we employ (mixture sets such that the closure condition on mixing asserts only the existence of mixtures on certain subsets of elements in the set, i.e. disjoint propositions and mixtures of them): the checking of the proof is completely straightforward, up to the assertion that is Theorem 7, and the subsequent definition of the utility u, in the original Herstein–Milnor paper (1953). These concluding steps of the Herstein–Milnor theorem require that for arbitrary r_0 and r_1 there are elements a and b such that r_0 and r_1 fall in S_{ab} (the subset of the ordered preferences lying between elements a and b), and such that mixtures on a and b are defined. The axiom that has been added, Axiom 7, guarantees that such disjoint elements a and b exist. The proof then goes through as in the original Herstein–Milnor paper.

So alter Fishburn's proof as follows: Where the antecedents of assertions in Fishburn's proof mention mixtures, add to the antecedents of the assertions that the mixtures are well-defined. Where the consequents of assertions in Fishburn's proof assert the existence of mixtures (for scaling the utility functions) on x and y, say $\alpha x + (1 - \alpha)y$, substitute assertion of the existence of $\alpha x^* + (1 - \alpha)y^*$, which is obtained as follows:

Suppose $x \geqslant y$. Apply Axiom 7 to find disjoint P and Q such that $P \geqslant x \geqslant y \geqslant Q$. Use the version of Herstein–Milnor that holds for the sets of mixtures on partitions to obtain

$$x^* = \beta_x P + (1 - \beta_x)Q \sim x, \quad \text{and}$$
$$y^* = \beta_y P + (1 - \beta_y)Q \sim y.$$

So $\alpha x^* + (1 - \alpha)y^*$ is well-defined (Axiom 2a), and is ranked with $\alpha x + (1 - \alpha)y$ (Herstein–Milnor for mixtures on partitions). The mixture $\alpha x^* + (1 - \alpha)y^*$ will play the role in scaling our system's utilities that the mixture $\alpha x + (1 - \alpha)y$ (which may not be well-defined) played in Fishburn's proof.

COMPARISON WITH JEFFREY–BOLKER; SPECULATION ABOUT ANOTHER FOUNDATION

The foundation presented above is, in some important ways, better than others in the literature: most important, it is a foundation for the best theory of rational choice available, and it generates that theory in a natural way. In other respects, this foundation fares no worse than others: the constraints on rational preference systems imposed by the axioms are at least as plausible as conditions for rational preference as are those of other theories; the richness assumptions about the preference orderings are strong, but no stronger than assumptions generally made by other theories. In one respect, however, another theory has an advantage this theory lacks: the Jeffrey–Bolker theory has, as I noted before, the very nice feature that lotteries or mixtures are not built into the preference ordering – so in that theory it is really true that probability and utility measures are derived from purely *qualitative* preference.[12] It is true that the Jeffrey–Bolker foundation also has unusual uniqueness results, weaker than those of other theories, but in my view this does not detract from the elegance of the theory. In the terminology of Krantz *et al.* (1971), the Jeffrey–Bolker theory provides a *fundamental measurement* of preference by probability and utility, whereas other theories do not.

Can a theory with this virtue of the Jeffrey–Bolker theory be given for conditional preference, generating causal decision theory in the natural way I have described above? I think the answer is probably "yes", though I do not have the theory now. I cannot say very much more than it is a project worth pursuing, but the following seems worth mentioning:

1. Since (a) the Jeffrey–Bolker axioms guarantee the existence of closely related pairs of probability and utility functions for any preference ordering satisfying the axioms, (b) the utility functions are order-preserving, and (c) the causal counterexamples to V-maximization show that the Jeffrey–Bolker utilities mismeasure the agent's preferences in those situations, it is clear that the agent's preferences violate one or more of the axioms of the Jeffrey–Bolker theory. It strikes me that among the axioms the only plausible candidate for a violation that accounts for the causal counterexamples is the Impartiality condition.[13] Getting clear about exactly how Impartiality is violated in these problems should be a useful step in the project.

2. Another axiom important to the Jeffrey–Bolker theory is their Averaging condition,[14] which is like an important axiom of Fishburn's (see Axiom 9). In both theories the axioms seem quite reasonable. Can their presence in the two theories be exploited?

These points seem worth noticing if it is the Jeffrey–Bolker theory that one seeks to adapt to get the kind of foundation under discussion. The other obvious place to look for ideas is the Luce and Krantz theory – in spite of their commitment to the peculiar disjunctive conditional preferences, that theory (or parts of it) may well be useful to the project. Of course, the idea of looking to Jeffrey–Bolker has been strongly suggested to me by the points Skyrms and Jeffrey have made about applying it to causal decision theory. As I mentioned early in the paper (see note 1), they have both sketched foundations for causal decision theory. In both cases they apply Jeffrey–Bolker theory in combination with an assumed prior specification of appropriate K-partitions. I would like to see a foundation as elegant as Jeffrey–Bolker that preserves the solution to the problem of selecting K-partitions I have presented above. Until one is given, I think the foundation presented here (based on Fishburn's theory, let me emphasize) is the best available.

Department of Philosophy
Ohio State University

NOTES

[1] It is the second of the virtues that I want to emphasize for this foundation: I am aware of two other sketches of foundations for causal decision theory which perhaps share the first virtue. Richard Jeffrey presents one in the concluding section of his (1981), and Brian Skyrms gives another in the concluding section of his (1982). Both suggestions depend on a prior specification of appropriate K-propositions (dependency hypotheses), and so they lack the second virtue described more fully below.

[2] The Twin Prisoners' Dilemma is a standard Prisoner's Dilemma, with the additional assumption that each prisoner believes that the other prisoner makes choices very much as he does. So each prisoner believes that there is a strong statistical correlation between his choice and the other's choice. For prisoner A, Pr(B rats / I rat) is high, and Pr(B cooperates / I cooperate) is high. Likewise for B. Each prisoner still believes, however, that his choice and action do not causally influence the other's choice and action. See Jeffrey (1983), Example 11 in Chapter 1.

[3] Suppose I value the pleasure I would derive from smoking this cigarette (act A_s) at 2 utiles, while I attach a large negative utility to contracting cancer, -1000 utiles. My utility function might say $V(S) = 2$, $V(-S) = 0$, $V(X) = -1000$, $V(-X) = 0$, where S is "I enjoy this cigarette," and X is "I contract cancer." Suppose I have these degrees of belief, where A_r is the act of refraining from smoking the cigarette:

$$\begin{array}{ll} P(S \;\&\; X/A_s) = 0.7 & P(S \;\&\; X/A_r) = 0.01 \\ P(S \;\&\; -X/A_s) = 0.25 & P(S \;\&\; -X/A_r) = 0.04 \\ P(-S \;\&\; X/A_s) = 0.04 & P(-S \;\&\; X/A_r) = 0.25 \\ P(-S \;\&\; -X/A_s) = 0.01 & P(-S \;\&\; -X/A_r) = 0.7 \end{array}$$

(J) tells us that

$$\begin{aligned} V(A_s) &= \sum_i P(C_i/A_s)\, V(C_i \;\&\; A_s) \\ &= 0.7(-998) + 0.25(2) + 0.04(-1000) + 0.01(0) \approx -738 \quad \text{and} \\ V(A_r) &= 0.01(-998) + 0.04(2) + 0.25(-1000) + 0.7(0) \approx -260 \end{aligned}$$

[4] See Stalnaker (1972) for the suggestion these theories build on.

[5] It may be that the agent is unsure about the causal structure of his decision problem:

he may have partial belief in a number of hypotheses about it. It is important to note that causal expected utility theory adequately handles these more complicated and more complicated and more realistic decision problems. The idea is simple: build his various hypotheses about the causal story into his appropriate K-partition. The hypotheses describe states of affairs beyond his influence which are relevant to the outcomes of his action. Each hypothesis will suggest a partition of factors which, according to the hypothesis, are appropriate K's; the expanded partition whose members are conjunctions of these factors with the hypotheses will be an appropriate K-partition. For details on this, see Skyrms (1979, pp. 136–138). The idea goes back to Savage (1954).

[6] I regard the sophisticated V-maximization theories as very interesting, but not because they are theories as good as causal decision theory. Only unreasonably strong assumptions about the agent's self-knowledge and about his fallibility in executing his chosen actions will save these V-maximization theories from *approximating* the correct answers that causal decision theory gives, or from getting the correct answers *almost* all the time. See Jeffrey (1981, 1983) and Eells (1982) for the variations.

[7] Eells (1982).

[8] In the following discussion I will sometimes use the phrase, "the agent has a preference for x" to mean simply "x appears in the agent's preference ranking," rather than "x ranks higher than some y in the agent's preference system.

[9] So like all such theories we fail here to provide a "fundamental measurement" of preference, since the mixing coefficients build some probability structure into the preferences from the beginning. I will discuss this point and the possibility of giving a mixture-free foundation below.

[10] A more extensive treatment of the notion of conditional preference appears in Armendt (1986b). An important point made there is the following: A preference P,Q is the agent's preference for P, under the supposition that Q holds. And the supposition that Q is one in which the agent's beliefs about the causal structure of the world are minimally altered (unless explicitly cancelled by Q). The supposition that Q is not a supposition that Q occurs no matter what the agent does, for example (though the agent might make such a supposition; the point here is that it is a *different* supposition from simply supposing that Q). I take this tendency to preserve the agent's beliefs about causal structures – the ways states, acts, and consequences produce each other in the actual world – as entirely appropriate in a theory of conditional preference tied to rational *action*, where the agent is interested in producing (in the actual world) desirable actions and consequences.

[11] In the precise statement of this rule given below, $P(B/B \vee C)$ is written $P_{B \vee C}(B)$. The representation theorem actually finds probability measures P_Q on subsets of $\mathscr{C}$ for every proposition Q in $\mathscr{C}'$. It establishes that these probabilities behave as conditional probabilities; see clause (iii) of the theorem. The probability measure P in the text is the measure P_T.

[12] The Luce and Krantz (1971) theory also has a purely qualitative preference ordering, but as I mentioned before, it incorporates disjunctive conditional preferences that I find very problematic.

[13] The Impartiality condition states that if A, B, and C are pairwise disjoint, and $A \sim B \sim C$, and $(A \vee C) \sim (B \vee C)$, then for all D disjoint from A and from B, $(A \vee D) \sim (B \vee D)$. See Jeffrey (1983), p. 147.

[14] Averaging says that if A and B are disjoint, and $A \succcurlyeq B$, then $A \succcurlyeq A \vee B \succcurlyeq B$. See Jeffrey (1983), p. 146.

REFERENCES

Armendt, B. (1983) *Rational Decision Theory: The Foundations of Causal Decision Theory*. Ph.D. dissertation for the Department of Philosophy, University of Illinois at Chicago.

Armendt, B. (1986a) 'A foundation for causal decision theory', *Topoi*, **5**, 3–19.

Armendt, B. (1986b) 'Conditional preference and rational choice.' Presented at the 1986 meetings of the Pacific Division of the American Philosophical Association.

Balch, M. and Fishburn, P. (1974) 'Subjective expected utility for conditional primitives', in *Essays on Economic Behavior under Uncertainty*, Balch, McFadden, and Wu (eds.), North-Holland.

Bolker, E. (1966) 'Functions resembling quotients of measures', *Transactions of American Mathematical Society* **124**, 292–312.

Bolker, E. (1967) 'A simultaneous axiomatization of utility and subjective probability', *Philosophy of Science* **34**, 333–340.

Eells, E. (1982) *Rational Decision and Causality*, Cambridge University Press.

Fishburn, P. (1973) 'A mixture-set axiomatization of conditional subjective expected utility', *Econometrica* **41**, 1–25.

Gibbard, A. and Harper, W. (1976) 'Counterfactuals and two kinds of expected utility', in *Ifs*, Harper, Stalnaker, and Pearce (eds.), Reidel.

Herstein, I. and Milnor, J. (1953) 'An axiomatic approach to measurable utility', *Econometrica* **21**, 291–297.

Jeffrey, R. (1965) *The Logic of Decision*, McGraw Hill.

Jeffrey, R. (1981) 'The logic of decision defended', *Synthese* **48**, 473–492.

Jeffrey, R. (1983) *The Logic of Decision*, 2nd edition, University of Chicago Press.

Krantz, D., Luce, R. D., Suppes, P., and Tversky, A. (1971) *Foundations of Measurement*, Vol. I, Academic Press.

Lewis, D. (1981) 'Causal decision theory', *Australasian Journal of Philosophy* **59**, 5–30.

Luce, R. D. and Krantz, D. (1971) 'Conditional expected utility', *Econometrica* **39**, 253–271.

Savage, L. J. (1954) *The Foundations of Statistics*, Wiley.

Skyrms, B. (1979) *Causal Necessity*, Yale University Press.

Skyrms, B. (1982) 'Causal decision theory', *Journal of Philosophy* **79**, 695–711.

Skyrms, B. (1984) *Pragmatics and Empiricism*, Yale University Press.

Sobel, J. H. (1978) *Probability, Chance, and Choice: A Theory of Rational Agency*, unpublished.

Stalnaker, R. (1972) 'Letter to David Lewis', in *Ifs*, Harper, Stalnaker, and Pearce (eds.), Reidel.

von Neumann, J. and Morgenstern, O. (1947) *Theory of Games and Economic Behavior*, 2nd edition, Princeton University Press.