(version of 2/1/03)

The Logical Foundations of Decision-Theoretic Planning in Autonomous Agents

John L. Pollock
Department of Philosophy
University of Arizona
Tucson, Arizona 85721
pollock@arizona.edu
http://www.u.arizona.edu/~pollock

Abstract

Decision-theoretic planning is normally based on the assumption that plans can be compared by comparing their expected-values, and the objective is to find an optimal plan. This is typically defended by reference to classical decision theory. However, classical decision theory is actually incompatible with this "simple plan-based decision theory". A defense of plan-based decision theory must begin by showing that classical decision theory is incorrect insofar as the two theories conflict, so this paper begins by raising objections to classical decision theory . First, there is a discussion of the considerations arising out of the Newcomb problem that have given rise to causal decision theory. Next, counterexamples are constructed for classical decision theory turning on the fact that an agent may be unable to perform an action, and may even be unable to try to perform an action. A proposal is made for how to repair classical decision theory in light of these counterexamples. But then turning to the concept of an "alternative" that is presupposed by classical decision theory, it is argued that actions must often be chosen in groups rather than individually, i.e., the objects of rational choice are plans. It is argued that optimality cannot be defined for plans, and even if it could be, it would not be reasonable to require rational agents to find optimal plans. So simple plan-based decision theory must also be rejected. An alternative called "locally global planning" is proposed as a replacement for both classical decision theory and simple plan-based decision theory.

1. Decision-Theoretic Planning

Planning by rational agents operating in environments of real-world complexity must be decision-theoretic. This general point is widely appreciated in the AI planning community, and there is a lot of current interest in constructing decision-theoretic planners (for example, see Blythe and Veloso (1997), Boutelier et al (1999), Haddawy and Hanks (1990), Ngo, Haddawy and Nguyen (1998), Onder and Pollack (1997, 1999), Onder, Pollack and Horty (1998), and Williamson and Hanks (1994).). However, the rush to implementation has proceeded without careful consideration of the theoretical foundations of such planning.

The normal presumption in decision-theoretic planning has been that the expected-value of a plan should be identified with the expected-value of executing the plan (performing the actions prescribed by the plan in the order prescribed) and plans can be compared directly in terms of

This work was supported by NSF grant no. IRI-IIS-0080888.

their expected-values. It is supposed that the objective of decision-theoretic planning is to find an optimal plan, i.e., a plan such that no alternative plan has a higher expected-value. This approach to decision-theoretic planning amounts to taking classical decision theory, which is about choosing between alternative actions, and applying it directly to plans, simply replacing "actions" by "plans" throughout the theory. It seems to be assumed that this is somehow justified by classical decision theory itself, although it is hard to see how that could be the case. Classical decision theory is about actions. It tells us to choose actions one at a time on the basis of their expected-values, not as a package (a plan) to be evaluated in combination. A consequence of adopting a plan is a decision to perform the actions prescribed by it. Evaluating plans decision-theoretically has the potential to conflict with classical decision theory by prescribing the performance of actions contained in plans when those actions would not be individually optimal (and so would not be chosen by classical decision theory). If this can happen, then rather than supporting plan-based decision theory, classical decision theory entails the falsity of plan-based decision theory. And as I will show below (see also my (1992) and (1995)), this can happen. It follows that proponents of plan-based decision theory cannot get away with just waving their hands vaguely at classical decision theory and saying "so we can evaluate plans decision-theoretically".

If plan-based decision theory is true, classical decision theory must be false. So a defense of plan-based decision theory must be based upon a criticism of classical decision theory. The general logic of my account in this paper is that classical decision theory is subject to several important theoretical difficulties, and the repair of these difficulties leads eventually to a properly formulated plan-based decision theory. I will discuss three such difficulties. The first is the set of familiar counterexamples associated with the "Newcomb problem" that have given rise to "causal decision theory". Second, I will raise objections to the classical definiton of "expected-value". This problem is fixable, but it is important to discuss it because the repair that is required must be applied to the use of expected-values in plan-based decision theory as well. Finally, I will present an argument to the effect that the proper objects of rational choice are plans rather than actions. Actions become derivatively rational by being prescribed by rationally chosen plans. However, I will also argue that plan-based decision theory cannot be viewed as a search for optimal plans. I will argue that optimality is not even well-defined outside of toy examples, so plan adoption must be based on somewhat different considerations. I will propose a theory that I call *locally global planning.*

2. Classical Decision Theory

Classical decision theory is a theory of rational choice. It is a theory of how an agent should, rationally, go about deciding what actions to perform at any given time. It is assumed that these decisions must be made in the face of uncertainty regarding both the agent's initial situation and the consequences of his actions. By "classical decision theory" I mean the nexus of ideas stemming in part from Ramsey (1926), von Neumann and Morgenstern (1944), Savage (1954), Jeffrey (1965), and others who have generalized and expanded upon it. The different formulations look very different, but the basic prescription of classical decision theory can be stated simply. We assume that our task is to choose an action from a set A of *alternative actions*. The actions are to be evaluated in terms of their outcomes. We assume that the *possible outcomes* of performing these actions are partitioned into a set O of pairwise exclusive and jointly exhaustive outcomes. We further assume that we know the probability $\textbf{PROB}(O/A)$ of each outcome conditional on the performance of each action. Finally, we assume a *utility-measure* $\mathbf{U}(O)$ assigning a numerical utility value to each possible outcome. The *expected value* of an action is defined to be a weighted average of the values of the outcomes, discounting each by the probability of that outcome occurring if the action is performed:

(1) $$\mathbf{EV}(A) = \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot \textbf{PROB}(O/A).$$

The crux of classical decision theory is that actions are to be compared in terms of their expected-values, and rationality dictates choosing an action that is *optimal*, i.e., such that no alternative has a higher expected-value.

More realistically, there may be an infinite set $\mathcal{W}$ of possible outcomes ("possible worlds"). The classical theory is extended to accommodate this possibility. If we have a probability distribution over the values of the worlds, we can generalize the definition of expected-value as follows:

(2) $$\mathbf{EV}(A) = \int_{-\infty}^{\infty} r \cdot \frac{d}{dr} \textbf{PROB}(\mathbf{U}(w) \leq r \,/\, w \in \mathcal{W} \;\&\; A \text{ is performed in } w)dr$$

where w is a random variable ranging over possible worlds in which A is performed. If there are just finitely many outcomes, (1) can be derived from (2). Technically, the integral in (2) characterizes the mathematical expectation of the function $\mathbf{U}(w)$ on the condition "$w \in \mathcal{W}$ & A is performed in w".

One point requires clarification. In saying that we are choosing from a set of alternative actions, the implication is that we will only perform one of them. But classical decision theory is mute on the question of whether we *could* perform (or try to perform) more than one. If we can, that will affect the expected-value of performing an action. That expected-value of A will factor in the possibility that we will also perform some of the other alternatives at the same time. That is not what we want, however. In talking about the expected-value of performing an action, what we mean to capture is the expected-value of performing the action and *not* any of the other alternative actions. So let us build that into the specification of the actions.

3. Causal Decision Theory

My first objection to classical decision theory is a familiar one that has figured prominently in the recent decision-theoretic literature, although it is not well-known in AI. Nozick's (1969) presentation of the Newcomb problem led to a general recognition that classical decision theory is flawed, making incorrect prescriptions in some cases. The Newcomb problem itself commands conflicting intuitions, but there are other examples that are clearer. One of the more compelling examples is due to Stalnaker (1978). You are deciding whether to smoke. Suppose you know that smoking is somewhat pleasurable, and harmless. However, there is also a "smoking gene" present in many people, and that gene both (1) causes them to desire to smoke and (2) predisposes them to get cancer (but not by smoking). Smoking is evidence that one has the smoking gene, and so it raises the probability that one will get cancer. Getting cancer more than outweighs the pleasure one will get from smoking, so classical decision theory recommends against smoking. But this seems clearly wrong. Smoking does not *cause* cancer. It is just evidence that one already has the smoking gene and hence may get cancer from that. If you have the smoking gene, you will still have it even if you refrain from smoking, so the latter will not prevent your getting cancer.

As a number of authors (Gibbard and Harper 1978; Sobel 1978; Skyrms 1980, 1982, 1984; Lewis 1981) have observed, conditional probabilities can reflect either evidential connections or causal connections. In this example, the connection between smoking and getting cancer is merely evidential. That is, smoking is evidence for cancer, but it does not cause it. In deciding whether to perform an action, we consider the consequences of performing it. The consequences should be its *causal consequences*, not its evidential consequences. This suggests that a correct formulation of decision theory should replace the conditional probability $\textbf{PROB}(O\,/\,A)$ by some kind of "causal probability" $\textbf{C-PROB}_A(O)$. The resulting theories are called *causal decision theories*.

A number of proposals have been made for how to define causal probability. Lewis (1981) argues that they are all closely related. In my (2002), I defended a theory having the same logical form as Skyrms' (1980, 1982, 1984) theory, but based on a somewhat different idea. In the present paper I will sketch my theory of causal probability, and I will use it throughout the paper, but to avoid getting bogged down in what is basically a complex side-issue, I will not give any detailed defense of my theory, other than pointing out that it handles the known counterexamples in a congenial way. For a more sustained defense, see my (2002).

3.1 Defining Causal Probability

The basic idea behind my proposal is simple — causal probability propogates forward in time, never backward. My suggestion is that in computing the possible effects of an action, we think of the world as evolving causally over time, interject the action into the world at the appropriate point, and then propogate changes forward in time as the world continues to evolve. This way of conceptualizing the world as evolving in temporal order is precisely the same idea that underlies most current solutions to the frame problem in AI (see Shoham (1986,1987), Hanks and McDermott (1986,1987), Lifschitz (1987), Gelfond and Lifschitz (1993), Shanahan (1990,1995,1996,1997), Pollock (1998)). Those solutions are based upon the idea that given a set of deterministic causal laws, to compute the result of a sequence of actions we imagine them occurring in temporal order and propogate the changes through the world in that order. As I will define it, causal probability does the same thing probabilistically.

To make this precise, let us begin with the simplifying assumption that actions occur instantaneously. They have dates that are single instants of time. These are *point-dated* actions. Singular states of affairs also have dates, but I will allow them to be either time intervals or time instants (degenerate intervals). I also assume that we can assign dates to logical combinations built out of conjunctions, disjunctions, and negations of singular states of affairs. The date of a negation is the date of what it negates, the date of a conjunction is the union of the dates of the conjuncts, and the date of a disjunction is the union of the dates of the disjuncts. The date of such a combination can be a time interval with gaps. I will refer to these logical combinations of singular states of affairs as *states of affairs* (dropping "singular"). Let us say that *Q postdates P* iff every time in the date (an interval) of *P* is < every time in the date of *Q*. Let us say that *P predates Q* iff every time in the date (an interval) of *P* is ≤ every time in the date of *Q*. So if a state predates a point-dated action, the end-point of its date may be the same as the date of the action. But if it postdates the action, it occurs wholly after the date of the action. I assume that an action cannot have a causal influence on a state that predates it.[1]

---

[1] For a defense of this in the case where the action and the state have the same point date, see the discussion of causal necessitation in Pollock (1998).

To see how to define causal probability, suppose first that the world is deterministic. This means that each complete state of the world determines each subsequent state. The determination is by physical laws. Each state nomically implies subsequent states, i.e., laws of nature entail that if the first state occurs it will be followed by the second. In asking whether a possible outcome would result from a particular world state in which an action is performed, we are asking whether the outcome will be present in subsequent states. In a deterministic world, O will result just in case the actual state of the world up to and including the time A is performed includes a set B of singular states of affairs such that $A\&B$ nomically implies O. I will call B a *background state* for O relative to A.

If we are uncertain about the precise state of the world, then we may be uncertain about whether O will result. The probability that O will result should be identified with the probability that the state of the world at the time A is performed contains a background state for O relative to A. If B is the only background state for O relative to A, then the probability of O given A should be identified with $\textbf{PROB}(B)$. If instead there are a finite number of such background states $B_1,\ldots,B_n$, then the probability of O given A should be identified with $\textbf{PROB}(B_1 \vee \ldots \vee B_n)$. Let us write this probability as $\textbf{C-PROB}_A(O)$.

$\textbf{C-PROB}_A(O)$ need not be the same as $\textbf{PROB}(O / A)$. The latter would be $\textbf{PROB}(B_1 \vee \ldots \vee B_n / A)$. A cannot *cause* changes to the background state, but it can be evidence regarding whether a background state is actual. This is precisely what happens in the smoking gene example. If we suppose that the gene causes cancer deterministically, then G is the sole background state and $\textbf{PROB}(G / A) \neq \textbf{PROB}(G)$. The probability $\textbf{C-PROB}_A(O)$ is then equal to $\textbf{PROB}(G)$ rather than $\textbf{PROB}(G / A)$. This is a *causal probability* that results from propogating the effects of actions forward in time but not backward in time. We hold the background state fixed, assigning to background states whatever probabilities they have prior to the action's being performed.

If we turn to nondeterministic worlds, the background states may no longer nomically imply the outcomes. They may only confer probabilities on the outcomes. If there were a single background B such that O can only result from A by having B true, we could define

$$\textbf{C-PROB}_A(O) = \textbf{PROB}(B) \cdot \textbf{PROB}(O / A\&B).$$

More generally, if we could confine our attention to a finite set B of (pairwise logically disjoint) backgrounds, we could define:

$$\textbf{C-PROB}_A(O) = \sum_{B \in \mathrm{B}} \textbf{PROB}(B) \cdot \textbf{PROB}(O / A\&B).$$

That is, the causal probability is the mathematical expectation of the probability of the outcome on the different possible backgrounds.

To define $\text{C-PROB}_A(O)$ generally (when O postdates A and A is a point-dated action), let C be the set of all singular states of affairs and negations of singular states of affairs predating A. Define an *A-world-state* to be a maximal subset of C nomically consistent with A. I will not usually distinguish between an A-world-state and the conjunction of its members. Let W be the set of all A-world-states. Then we can define $\text{C-PROB}_A(O)$ to be the mathematical expectation of the probability of the outcome on the different possible A-world-states. If W is finite, our definition becomes:

$$\text{C-PROB}_A(O) = \sum_{W \in \mathrm{W}} \text{PROB}(W) \cdot \text{PROB}(O / A \& W).$$

Realistically, W will be infinite, in which case $\text{C-PROB}_A(O)$ must be defined using the integral definition of expected-value:

$$\text{C-PROB}_A(O) = \int_{-\infty}^{\infty} r \cdot \frac{\mathrm{d}}{\mathrm{d}r} \text{PROB}(\text{PROB}(O / A \& W) \le r)\, \mathrm{d}r$$

where W is a random variable ranging over members of W. However, to keep the mathematics simple, I will pretend that W is finite and use the summation version of the definition. This will make no difference to the results. It is easily verified that C-PROB_A is a probability, i.e., that it satisfies the probability calculus.

The fundamental idea behind this definition of causal probability is that in computing how likely an outcome is to result from an action, we want to propogate changes forward in time rather than backward. A useful way of conceptualizing this is to think of the world as described by different *scenarios*, each consisting of some A-world-state being true, followed by the action, followed by an outcome. The scenarios can be diagrammed in the form of a tree, as in figure 1. Scenarios depict the world as “evolving in temporal order”. The causal effects of events propogate forward in time, and actions only affect the probabilities of events occurring after them. A probability can be associated with a scenario by starting with the start state and computing the probability of each subsequent state in the scenario in terms of the probabilities of actions producing those states given the earlier states of the scenario. In other words, the probabilistic consequences of events are propogated forwards in time. For a scenario passing through W and terminating with O, the associated probability is $\text{PROB}(W)\cdot\text{PROB}(O / A\&W)$. The causal probability $\text{C-PROB}_A(O)$ is the sum of the probabilities of the scenarios terminating with O.

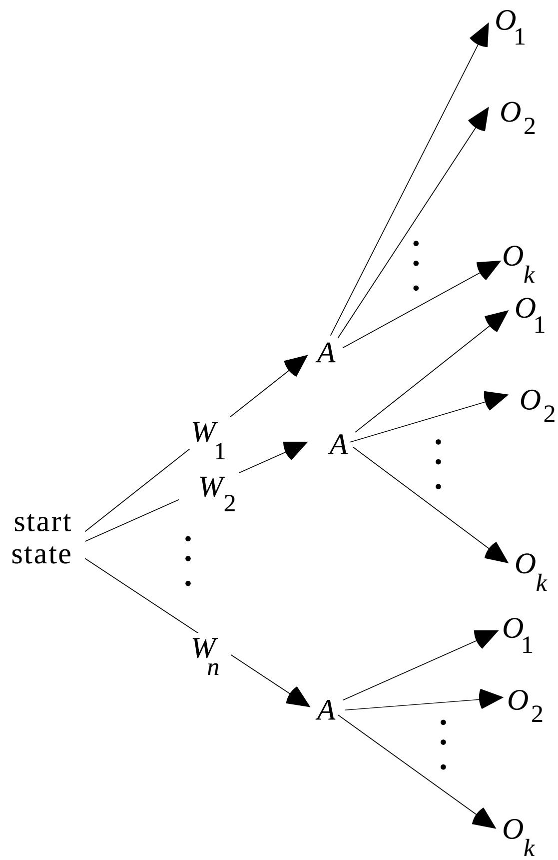

Figure 1. *S*cenarios evolving with the passage of time

3.2 Computing Causal Probability

If causal probability is to be useful, there must be efficient ways of computing it. If we had to compute $\text{C-PROB}_A(O)$ by actually performing the summation (or integration) involved in the definition, the task would be formidable. Fortunately, this computation can be simplified considerably. Recall that $\mathcal{C}$ is the set of "constituents" of A-world-states. Let us say that a subset S of $\mathcal{C}$ *shadows* A with respect to O iff (1) S is nomically consistent with A, (2) for every $W \in \mathcal{W}$ and any S^{**}, if $S \subseteq S^{**} \subseteq W$ then $\text{PROB}(O / A\&S^{**}) = \text{PROB}(O / A\&S)$, and (3) there is no proper subset S^* of S such that for every $W \in \mathcal{W}$ and any S^{**}, if $S^* \subseteq S^{**} \subseteq W$ then $\text{PROB}(O / A\&S^{**}) = \text{PROB}(O / A\&S)$. The shadows are minimal descriptions of all aspects of the A-world-state relevant to the evaluation of the probability of O. Let $\mathcal{S}$ be the set of all shadows. Shadows can be constructed by starting from members of $\mathcal{W}$ and then removing elements that do not affect the probability of O. It follows that every A-world-state W contains a shadow S such that $\text{PROB}(O / A\&W) = \text{PROB}(O / A\&S)$.

Let C* be the set of all members of C occurring in one or more of the shadows. Define a *background* to be a maximal subset of C* nomically consistent with *A*. Let B be the set of all backgrounds. The backgrounds form a partition. That is, they are pairwise logically disjoint and the disjunction of all of them is a (nomically) necessary truth. Now suppose $B \in \mathrm{B}$ is a background and $W \in \mathrm{W}$ and $B \subseteq W$. W contains a shadow S such that $\mathbf{PROB}(O / A\&W) = \mathbf{PROB}(O / A\&S)$, and the shadow consists of members of C*, so $S \subseteq B$, and hence by the definition of "shadow", $\mathbf{PROB}(O / A\&B) = \mathbf{PROB}(O / A\&S)$. Thus $\mathbf{PROB}(O / A\&W) = \mathbf{PROB}(O / A\&B)$. This enables us to prove a central theorem in the theory of causal probability:

Theorem 1: $\mathbf{C\text{-}PROB}_A(O) = \sum_{B \in \mathrm{B}} \mathbf{PROB}(B) \cdot \mathbf{PROB}(O / A\&B)$.

(The proof of this and other theorems stated without proof can be found in my (2002).)

Once causal probability has been defined, causal decision theory proposes to redefine the expected-value of an action in terms of causal probabilities:

$$\mathbf{EV}(A) = \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot \mathbf{C\text{-}PROB}_A(O).$$

The case of infinitely many outcomes is handled as before, replacing **PROB** by **C-PROB**.

Let us see how this proposal handles the smoking gene example. If a person has the smoking gene (*G*), he already has it when he makes his decision whether to smoke (*S*). If we suppose that the only part of an *S*-world-state that makes any difference to the probability of getting cancer is *G* or ~*G*, it follows that the set of shadows S = {{*G*},{~*G*}}, and so the set of backgrounds B = {{*G*},{~*G*}}. Hence

$$\mathbf{C\text{-}PROB}_S(cancer) = \mathbf{PROB}(G) \cdot \mathbf{PROB}(cancer / S\&G) + \mathbf{PROB}(\sim G) \cdot \mathbf{PROB}(cancer / S\&\sim G).$$

Then the expected-value of smoking is computed as follows:

$$\mathbf{EV}(S) = \mathbf{U}(pleasure\ of\ smoking) \cdot \mathbf{C\text{-}PROB}_S(pleasure\ of\ smoking) + \mathbf{U}(cancer) \cdot \mathbf{C\text{-}PROB}_S(cancer)$$

$$= \mathbf{U}(pleasure\ of\ smoking) \cdot \mathbf{PROB}(pleasure\ of\ smoking / S) + \mathbf{U}(cancer) \cdot [\mathbf{PROB}(G) \cdot \mathbf{PROB}(cancer / S\&G) + \mathbf{PROB}(\sim G) \cdot \mathbf{PROB}(cancer / S\&\sim G)].$$

If we make the assumption that $\mathbf{PROB}(cancer / S\&G) = \mathbf{PROB}(cancer / G)$ and $\mathbf{PROB}(cancer / S\&\sim G) = \mathbf{PROB}(cancer / \sim G)$, it follows that:

$$\mathbf{EV}(S) = \mathbf{U}(\textit{pleasure of smoking})\cdot\textsf{PROB}(\textit{pleasure of smoking} / S) + \mathbf{U}(\textit{cancer})\cdot[\textsf{PROB}(G)\cdot\textsf{PROB}(\textit{cancer} / G) + \textsf{PROB}(\sim G)\cdot\textsf{PROB}(\textit{cancer} / \sim G)].$$

Similarly, the expected-value of not smoking ($\overline{S}$) is:

$$\mathbf{EV}(\overline{S}) = \mathbf{U}(\textit{cancer})\cdot[\textsf{PROB}(G)\cdot\textsf{PROB}(\textit{cancer} / G) + \textsf{PROB}(\sim G)\cdot\textsf{PROB}(\textit{cancer} / \sim G)].$$

Thus if $\mathbf{U}(\textit{pleasure of smoking}) > 0$ and $\textsf{PROB}(\textit{pleasure of smoking} / S) > 0$, it follows that $\mathbf{EV}(S) > \mathbf{EV}(\overline{S})$. So causal decision theory recommends smoking, which is the right choice.

Of course, realistically, other elements of *S*-world-states will also be statistically relevant to getting cancer, e.g., whether one's parents had the smoking gene. However, the effect of one's parents having the smoking gene is "screened off" by knowing whether one has the gene oneself, i.e., if you know whether you have the smoking gene, the additional knowledge of whether your parents had it does not effect the probability of getting cancer. So the set of shadows, and hence the set of backgrounds, remains unchanged.

Normally, shadows will be more numerous than in the smoking gene example. However, the shadows may not all be relevant. The need for causal probabilities only arises when the action is statistically relevant to some of the backgrounds. If the backgrounds are all statistically independent of the action, then the causal probability is the same as the classical probability:

Theorem 2: If for each $B \in \mathrm{B}$, $\textsf{PROB}(B / A) = \textsf{PROB}(B)$, then $\textsf{C-PROB}_A(O) = \textsf{PROB}(O / A)$.

More generally, the action may be statistically relevant to just a few constituents of the backgrounds. Then we can often make use of the following theorem:

Theorem 3: If $\mathrm{C}_0 \subseteq \mathrm{C}^*$, let B_0 be the set of all maximal subsets of C_0 nomically consistent with *A*, and let B^* be the set of all maximal subsets of $\mathrm{C}^* - \mathrm{C}_0$ nomically consistent with *A*. If for every $B_0 \in \mathrm{B}_0$ and $B^* \in \mathrm{B}^*$, $\textsf{PROB}(B^* / B_0 \& A) = \textsf{PROB}(B^* / B_0)$, then

$$\textsf{C-PROB}_A(O) = \sum_{B \in \mathrm{B}_0} \textsf{PROB}(B_0)\cdot\textsf{PROB}(O / A \& B_0).$$

So if there is a subset C_0 of constituents of backgrounds relative to which all other combinations of constituents are statistically independent of *A*, then we can compute causal probabilities by making reference only to backgrounds built out of the members of C_0. For example, suppose there are two constituents of backgrounds that are statistically relevant to getting cancer — having the smoking gene, and having been raised on a nuclear waste dump (*N*). Then B =

$\{\{G,N\},\{G,\sim N\},\{\sim N,G\},\{\sim N,\sim G\}\}$. However, S is not statistically relevant to whether one was raised on a nuclear waste dump, even given that one does or does not have the smoking gene:

$$\mathbf{PROB}(N / S\&G) = \mathbf{PROB}(N / G)$$
$$\mathbf{PROB}(N / S\&\sim G) = \mathbf{PROB}(N / \sim G)$$
$$\mathbf{PROB}(\sim N / S\&G) = \mathbf{PROB}(\sim N / G)$$
$$\mathbf{PROB}(\sim N / S\&\sim G) = \mathbf{PROB}(\sim N / \sim G)$$

So we can let $C_0 = \{\{G\},\{\sim G\}\}$, and once more compute $\mathbf{C\text{-}PROB}_S(cancer)$ by reference to the small set of backgrounds $B_0 = \{\{G\},\{\sim G\}\}$.

3.3 Extending the Definition

The upshot of these results is that causal probabilities will usually be computable by performing manageably small sums. In cases in which actions are statistically relevant to their backgrounds, **C-PROB**'s may be significantly easier to compute than **PROB**'s. **C-PROB**'s can be computed recursively by propogating probabilities forwards through scenarios. But if a later state can affect the **PROB** of an earlier state, then **PROB**'s cannot similarly be computed recursively. So for practical purposes, **C-PROB**'s are simpler than **PROB**'s. This suggests that instead of expressing theorem 2 by saying that causal probabilities usually behave classically, it might be better to say that classical probabilities usually behave causally.

Thus far, $\mathbf{C\text{-}PROB}_A(O)$ has been defined for all states of affairs postdating A. It will be convenient to define $\mathbf{C\text{-}PROB}_A(O)$ for a broader class of states of affairs, including states of affairs that do not postdate A. If O predates A we can stipulate:

$$\mathbf{C\text{-}PROB}_A(O) = \mathbf{PROB}(O).$$

If O_1 postdates A and O_2 predates A, then we will further stipulate that

$$\mathbf{C\text{-}PROB}_A(O_1 \& O_2) = \mathbf{PROB}(O_1)\cdot\mathbf{C\text{-}PROB}_A(O_2 / O_1).$$

I am making the simplifying assumption that actions occur instantaneously, and so their dates are time points rather than intervals. If a state of affairs neither predates A nor postdates A then its date must be an interval (possibly with gaps) with the date of A lying within the interval. I assume that such a state of affairs can be split into a “first part” predating A and a “second part” postdating A, and then the state of affairs can be represented as the conjunction of these two parts. This has the consequence that $\textsf{\textbf{C-PROB}}_A(O)$ is defined for all states of affairs O.

The preceding definition is in terms of conditional causal probabilities, which we have yet to define. The standard definition would be:

$$\textsf{\textbf{C-PROB}}_A(O\,/\,P) = \textsf{\textbf{C-PROB}}_A(O\&P)\,/\,\textsf{\textbf{C-PROB}}_A(P).$$

However, when P predates A, we defined $\textsf{\textbf{C-PROB}}_A(O\&P)$ in terms of $\textsf{\textbf{C-PROB}}_A(O\,/\,P)$, so we must find an independent definition for the latter. That can be done as follows (when P predates A):

$$\textsf{\textbf{C-PROB}}_A(O\,/\,P) = \sum_{W\in \mathrm{W}} \textsf{\textbf{PROB}}(W\,/\,P)\cdot\textsf{\textbf{PROB}}(O\,/\,A\&W\&P).$$

This definition can be recast in terms of backgrounds, just as in theorem 1:

Theorem 4: If P predates A then where B is the set of backgrounds for O relative to A that are consistent with P:

$$\textsf{\textbf{C-PROB}}_A(O\,/\,P) = \sum_{B\in \mathrm{B}} \textsf{\textbf{PROB}}(B\,/\,P)\cdot\textsf{\textbf{PROB}}(O\,/\,A\&B\&P).$$

Once conditional probabilities are defined as above for the case in which P predates A, non-conditional causal probabilities are defined in general (for point-dated actions), and so for all other cases we can stipulate conventionally that:

$$\textsf{\textbf{C-PROB}}_A(O\,/\,P) = \textsf{\textbf{C-PROB}}_A(O\&P)\,/\,\textsf{\textbf{C-PROB}}_A(P).$$

Just as for nonconditional probabilities, the conditional causal probabilities of states predating A behave classically:

Theorem 5: If P and Q predate A, $\textsf{\textbf{C-PROB}}_A(Q\,/\,P) = \textsf{\textbf{PROB}}(Q\,/\,P)$.

I have been making the simplifying assumption that actions are instantaneous. Realistically, actions take time, so their dates should be intervals rather than instants. However, the extention of causal probability to temporally extended actions is extremely complicated. This is discussed in my (2002). As long as the actions, backgrounds, and outcomes do not have overlapping dates, we can treat causal probabilities just as if the actions were instantaneous, but when these dates overlap the calculations must be modified. However, I will not pursue the details of that here.

In AI, the best known probabilistic planner is Buridan (Kushmerick, Hanks, and Weld 1995). The probability semantics for Buridan is in some ways reminiscent of causal probability, because it defines the probabilities of outcomes by propogating probabilities forwards in time in much the same way causal probability does. The probability semantics of Buridan is "congenial to" causal probability, however it is indeterminate whether Buridan is actually computing causal probabilities. In Buridan, actions are encoded as triples $\langle\tau,\rho,e\rangle$ where τ is the *trigger* and ρ is the probability that effect e will result from performing the action when the trigger is true. If triggers are or contain backgrounds, then the probabilities defined by the semantics are causal probabilities. If no trigger contains a background then the probabilities are classical. If some triggers contain backgrounds but other do not, the result is a kind of hybrid between classical and causal probabilities that probably has no intuitive significance. For example, in the smoking gene example, if the only trigger is the tautology $\top$ then the probability computed is the classical probability of getting cancer given that one smokes. If the triggers are G and $\sim G$ then the probability is the causal probability (which, as we have seen, is quite different). I suspect that Kushmerick *et al* intended to be computing classical probabilities, but this shows that their semantics is incorrect if viewed in that way. It is not correct for causal probabilities either unless we make the further stipulation that the triggers are or contain backgrounds.

4. Action Omnipotence

The problems leading to causal decision theory are familiar to decision-theorists, so for present purposes I will regard causal decision theory as the preferred variant of classical decision theory. Henceforth when I speak of classical decision theory, I mean this causal variant. I turn next to difficulties that are not so familiar. The simplest of these difficulties is that classical decision theory assumes that actions can be performed infallibly — *action omnipotence.* To see that this is indeed an assumption of classical decision theory, consider a simple counterexample based on the failure of the assumption. Suppose I offer you the following choice (to be made now). I will give you ten dollars if you wiggle your left index finger in ten minutes, but I will give you one hundred dollars if you wiggle your right index finger in ten minutes. The hitch is that your right index finger is currently paralyzed and you are unsure whether the paralysis will have worn off in ten minutes. Your assessment of how likely you are to be able to wiggle your right index finger in ten minutes is surely relevant to your rational decision, but classical decision theory makes no provision for this. It dictates instead that you should choose to wiggle your right index finger in ten minutes, even if it is improbable that you will be able to do that.

Classical decision theory prescribes this because it only looks at the probabilities that outcomes will result if actions are performed, and does not take account of how likely it is that the agent will be able to perform the action.

There are two ways we might try to repair classical decision theory to avoid this counterexample. One way is to restrict the scope of classical decision theory to actions that can be performed infallibly. The other is to modify the way in which actions are assessed. I will consider the former strategy first, and when it fails I will take that to motivate the second strategy.

4.1 Restricting the Scope of Decision Theory

Classical decision theory would only issue rational prescriptions if it were applied to actions that can be performed infallibly. Are there any? It is clear that most ordinary actions, like *make a cup of coffee*, or even *walk across the room* can go awry. One can try to perform them and fail. As we have seen, even very low level actions like *wiggle your finger* can fail. This strongly suggests that there are no actions for which action omnipotence holds.

However, there is one candidate that calls for further discussion. At one time I thought that although we cannot always perform an action, we can always try, and so classical decision theory should be restricted to tryings. Unfortunately, it is not true that we can always try. Suppose I show you a wooden block, then throw it in an incinerator where it is consumed, and then I ask you to paint it red. You not only cannot paint it red — you cannot even try to do so. There is nothing you could do that would count as trying.

In this example, the state of the world makes it impossible for you to paint the block red. But what makes it impossible for you to try to paint it red is not the state of the world but rather your beliefs about the state of the world. For example, suppose I fooled you and you just *think* I threw the block in the incinerator. Then although the action of painting the block red is one that someone else could perform, *you* cannot even try to do it. Conversely, if I did destroy the block but you do not believe that I did, you would not be able to paint it but you might be able to try. For instance, if you believe (incorrectly) that the block is at the focus of a set of paint sprayers activated by a switch, you could try to paint the block by throwing the switch. These examples illustrate that what you can try to do is affected by your beliefs, not just by the state of the world.

These considerations have convinced me that there are no actions that can be performed infallibly. Action omnipotence fails in two ways — we may fail to perform an action when we try, and we may not even be able to try. Hence there is no way to restrict classical decision theory to a privileged set of actions for which it yields the correct prescriptions. The only course left is to modify the way classical decision theory evaluates actions. This amounts to modifying the definition of “expected-value”. Let us explore this possibility.

4.2 Expected-Utility

It is tempting to suppose that we can handle the failure of action omnipotence by discounting the expected-value of performing an action by the probability that we will be able to perform it if we try:

$$\mathbf{EV}(A)\cdot\mathbf{PROB}(A\,/\,try\text{-}A).$$

That does not work, however. The values we must take account of in assessing an action include (1) the values of any goals achieved by performing the action, (2) execution costs, and (3) side-effects that are not among the goals or normal execution costs but are fortuitous consequences of performing or trying to perform the action under the present circumstances. The goals will presumably be consequences of successfully performing the action, but execution costs and side effects can result either from successfully performing the action or from just trying to perform it. For example, if I try unsuccessfully to move a boulder with a pry bar, I may expend a great deal of energy, and I might even pull a muscle in my back. These are execution costs and side effects, but they attach to the trying — not just to the successful doing. To include all of these values and costs in our assessment of the action, we must look at the expected-value of trying to perform the action rather than the expected-value of actually performing it:

$$\mathbf{EV}(try\text{-}A).$$

This will have the automatic effect of discounting costs and values attached to successfully performing the action by the probability that we will be able to perform it if we try, but it also factors in costs and values associated directly with trying.

A surprising qualification is required, however. Rather than looking at the expected-value of trying to perform an action *A*, I will argue that we must consider the conditional expected-value of trying to perform *A* *given that* the agent can try to perform *A*. Conditional expected-values are defined by conditionalizing the probabilities on the condition:

$$\mathbf{EV}(A\,/\,P) = \sum_{O\in\mathrm{O}} \mathbf{U}(O)\cdot\mathbf{PROB}(O\,/\,A\&P).$$

This is the expected-value of the action given the assumption that *P* is true. In non-causal classical decision theory, conditionalizing on the agent being able to try to perform *A* makes no difference, because the conditional expected-value of trying to perform *A* given that the agent can try to perform A is defined in terms of PROB(O / *the agent tries to perform A* & *the agent can try to perform A*). If the agent tries to perform A, it follows logically that he can try to perform A, so this probability is the same as PROB(O / *the agent tries to perform A*), and hence the conditional expected-value is the same as the unconditional expected-value. However, in causal decision theory, this equivalence fails. By definition,

$$\text{C-PROB}_{try\text{-}A}(O/\textit{the agent can try to perform } A)$$

$$= \sum_{B\in \mathrm{B}} [\text{PROB}(B/\textit{the agent can try to perform } A) \cdot \text{PROB}(O/B \,\&\, \textit{the agent tries to perform } A \,\&\, \textit{the agent can try to perform } A)].$$

As before,

$$\text{PROB}(O/B \,\&\, \textit{the agent tries to perform } A \,\&\, \textit{the agent can try to perform } A)$$

$$= \text{PROB}(O/B \,\&\, \textit{the agent tries to perform } A)$$

but it need not be the case that PROB(B /*the agent can try to perform A*) = PROB(B). Knowing that the agent can try to perform A may alter the probability of relevant earlier states of affairs. To illustrate, think of a war game in which we are considering using a certain airplane to attack our opponent. However, the airplane is parked on an undefended airstrip and there is the possibility that our opponent will destroy the airplane before we can use it, in which case we will not even be able to try to use it. Now consider the causal probabilities in this example. If the plane has not been destroyed by the time we plan to use it (a necessary condition for our being able to try to attack the enemy with it), that may increase the probability that the enemy's air force has been crippled by earlier attacks, and the probability of the new attack being successful will thereby be increased because the enemy will be less likely to be able to repel it. That is,

$$\text{PROB}(\textit{the enemy air force was crippled earlier/we can try to attack with the airplane})$$

$$> \text{PROB}(\textit{the enemy air force was crippled earlier})$$

with the result that

$\textbf{C-PROB}_{\textit{we try to attack with the airplane}}$(*the attack will be successful* / *we can try to attack*)

> $\textbf{C-PROB}_{\textit{we try to attack with the airplane}}$(*the attack will be successful*).

This raises the expected-value of trying to attack with the airplane given that one can try (i.e., given that the airplane is still there). Given this difference, it seems clear that it is the conditional causal probability that should be employed in evaluating the attempt to attack. In general, an action should be evaluated in terms of the conditional expected-value of trying to perform it given that one can try to perform it:

EV(*try-A* / *can-try-A*).

However, more is relevant to the assessment of an action than the conditional expected-value of trying to perform it given that one can try to perform it. As we have seen, we may not be able to try to perform an action. It seems apparent that in comparing two actions, if we know that we cannot try to perform one of them, then it should not be a contender in the choice. But more generally, we may be uncertain whether we will be able to try to perform the action at the appropriate time. For example, in the war game, the probability of the plane's being destroyed should affect our assessment of the action when we compare it with alternative ways of attacking our opponent. The obvious suggestion is that we should discount the conditional expected-value of trying to perform an action by the probability that, under the present circumstances, we can try to perform it:

EV(*try-A* / *can-try-A*)·**PROB**(*can-try-A*).

That does not quite work, however. Suppose you are in a situation in which you get at least ten dollars no matter what you do. You get another ten dollars if you do *A*, but you have only a 50% chance of being able to try to do *A*. If you can try to do *A*, you have a 100% chance of succeeding if you try. If, instead of doing *A*, you do *B*, you will get one dollar in addition to the ten dollars you get no matter what. Suppose you are guaranteed of being able to try to do *B*, and of doing *B* if you try. Given a choice between *A* and *B*, surely you should choose *A*. You have a 50% chance of being able to try to do *A*, and if you do try you will get ten extra dollars. You have a 100% chance of being able to try to do *B*, but if you try to do *B* you will only get one extra dollar.

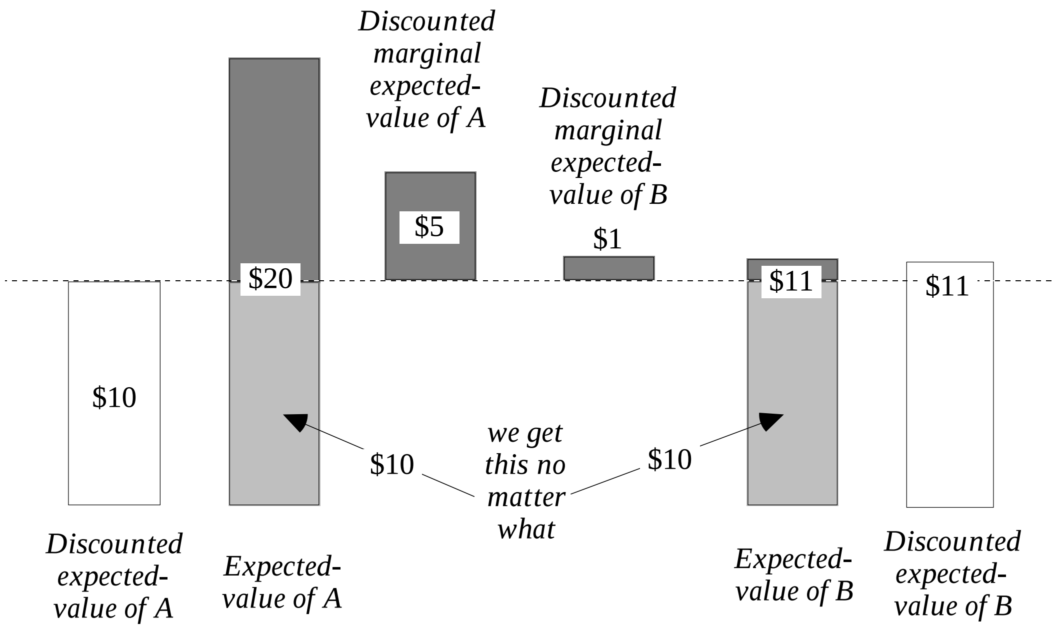

Figure 2. Discounted marginal expected-values.

However, the only possible outcome of trying to do *A* is worth twenty dollars, and the only possible outcome of trying to do *B* is worth eleven dollars. So these are the conditional expected-values of trying to perform *A* and *B*. If we discount each by the probability of being able to try to perform the action, the value for *A* is ten dollars and that for *B* is eleven dollars. This yields the wrong comparison. It is obvious what has gone wrong. We should not be comparing the total amounts we will get if we perform the actions, and discounting those by the probabilities of being able to try to perform the actions. Rather, we should be comparing the *extra amounts* we will get, over and above the ten dollars we will get no matter what, and discounting those extra amounts by the probabilities of being able to perform the actions. This is diagrammed in figure 2.

Apparently, in choosing between alternative actions, we must look at the expected-values that will accrue specifically as a result of performing each action. This is not the same thing as the expected-value of the outcome of the action, because some of the value of the outcome may be there regardless of what action is performed. The expected-value that will accrue specifically as a result of trying to perform *A* is the difference between the conditional expected-value of trying to perform *A* and the conditional expected-value of not trying to perform any of the alternative actions (*nil* — the "null-action"). This is the *marginal expected-value* of trying to perform the action. It is this marginal expected-value that should be discounted by the probability of being able to try to perform the action.

Putting this all together, I propose that we define the *expected-utility* of an action to be the marginal expected-value of trying to perform that action, discounted by the probability that we can try to do that:

$$\textbf{expected-utility}(A) = \text{PROB}(\textit{can-try-A})\cdot[\textbf{EV}(\textit{try-A} / \textit{can-try-A}) - \textbf{EV}(\textit{nil} / \textit{can-try-A})].$$

If the probability that the agent can try to perform A is 0, the conditional expected-value of trying to perform A is undefined, but in that case let us just stipulate that **expected-utility**(A) = 0. In classical decision theory, the terms "the expected-value of an action" and "the expected-utility of an action" are generally used interchangeably, but I am now making a distinction between them. My proposal is that we modify classical decision theory by taking it to dictate choosing between alternative actions on the basis of their expected-utilities. With this change, decision theory is able to handle the above examples in a reasonable way, without restricting it to a special class of particularly well-behaved actions.

This definition of the expected-utility of an action may seem a bit *ad hoc*. It was propounded to yield the right answer in decision problems, but why is this the right concept to use in evaluating actions? It will be shown in the next section that it has a simple intuitive significance. Comparing actions in terms of their expected-utilities is equivalent to comparing the expected-values of "conditional policies" of the form *try to do A if you can try to do A*.

5. Conditional Policies and Expected-Utilities

Decision theory has usually focused on choosing between alternative actions available to us *here and now*. A generalization of this problem will be important in understanding my proposal for reformulating decision theory to avoid the problems stemming from the failure of action omnipotence. We sometimes make *conditional decisions* about what to do if some condition P turns out to be true. For instance, I might deliberate about what route to take to my destination if I encounter road construction on my normal route. Where P predates A, *doing A if P* (and performing none of the alternative actions otherwise) is a *conditional policy*. Conditional decisions are choices between conditional policies. We can regard decision theory as telling us to make such conditional decisions on the basis of the expected-values of the conditional policies. For this purpose we must define the probability of an outcome conditional on a conditional policy:

$$\text{C-PROB}_{A\ if\ P}(O) = \text{PROB}(P)\cdot\text{C-PROB}_{A}(O / P) + \text{PROB}({\sim}P)\cdot\text{C-PROB}_{nil}(O/{\sim}P).$$

Then we can define straightforwardly:

$$\mathbf{EV}(A \textit{ if } P) = \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot \textsf{C-PROB}_{A \textit{ if } P}(O).$$

It is important to distinguish between the expected-value of a conditional policy and a conditional expected-value. The latter was defined as follows:

$$\mathbf{EV}(A / P) = \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot \textsf{C-PROB}_{A}(O / P).$$

This is the expected-value of the action given the assumption that P is true. **EV**(A *if* P), on the other hand, is the expected-value of doing A if P and doing nothing otherwise. The expected-value of a conditional policy is related to conditional expected-values as follows:

Theorem 6: $\mathbf{EV}(A \textit{ if } P) = \textsf{PROB}(P) \cdot \mathbf{EV}(A / P) + \textsf{PROB}({\sim}P) \cdot \mathbf{EV}(nil / {\sim}P)$.

Decision theory normally concerns itself with expected-values. However, the expected-value of an action is defined to be the expected-value of “the world” when the action is performed. This includes values that would have been achieved with or without the action. As we saw in section four, it is often more useful to talk about the marginal expected-value, which is the difference between the expected-value of the action and the expected-value of doing nothing. The marginal expected-value of an action measures how much value the action can be expected to *add* to the world:

$$\mathbf{MEV}(A) = \mathbf{EV}(A) - \mathbf{EV}(nil).$$

We can define conditional marginal expected-values and the marginal expected-values of conditional policies analogously:

$$\mathbf{MEV}(A / P) = \mathbf{EV}(A / P) - \mathbf{EV}(nil / P).$$

$$\mathbf{MEV}(A \textit{ if } P) = \mathbf{EV}(A \textit{ if } P) - \mathbf{EV}(nil \textit{ if } P).$$

The conditional policy *nil if P* prescribes doing *nil* if P and *nil* if ~P, so it is equivalent to *nil* simpliciter. Thus we could just as well have defined:

$$\mathbf{MEV}(A \textit{ if } P) = \mathbf{EV}(A \textit{ if } P) - \mathbf{EV}(nil).$$

It follows that comparing conditional policies in terms of their marginal expected-values is equivalent to comparing them in terms of their expected-values:

Theorem 7: $\mathbf{MEV}(A \textit{ if } P) > \mathbf{MEV}(B \textit{ if } Q)$ iff $\mathbf{EV}(A \textit{ if } P) > \mathbf{EV}(B \textit{ if } Q)$.

There is a simple relationship between the marginal expected-value of a conditional policy and the conditional marginal expected-value:

Theorem 8: $\mathbf{MEV}(A \textit{ if } P) = \mathsf{PROB}(P)\cdot\mathbf{MEV}(A / P)$

Proof:

$\mathbf{MEV}(A \textit{ if } P)$

$= \mathbf{EV}(A \textit{ if } P) - \mathbf{EV}(\textit{nil if } P)$

$= \mathsf{PROB}(P)\cdot\mathbf{EV}(A / P) + \mathsf{PROB}(\sim P)\cdot\mathbf{EV}(\textit{nil} / \sim P)$

$- \mathsf{PROB}(P)\cdot\mathbf{EV}(\textit{nil} / P) - \mathsf{PROB}(\sim P)\cdot\mathbf{EV}(\textit{nil} / \sim P)$

$= \mathsf{PROB}(P)\cdot\mathbf{EV}(A / P) - \mathsf{PROB}(P)\cdot\mathbf{EV}(\textit{nil} / P)$

$= \mathsf{PROB}(P)\cdot[\mathbf{EV}(A / P) - \mathbf{EV}(\textit{nil} / P)]$

$= \mathsf{PROB}(P)\cdot\mathbf{MEV}(A / P)$. ■

We defined **expected-utility**$(A) = \mathsf{PROB}(\textit{can-try-A})\cdot\mathbf{MEV}(\textit{try-A} / \textit{can-try-A})$, so by theorem 6:

Theorem 9: expected-utility$(A) = \mathbf{MEV}(\textit{try-A if can-try-A})$.

Hence by virtue of theorem 7:

expected-utility$(A) >$ **expected-utility**(B) iff $\mathbf{EV}(\textit{try-A if can-try-A}) > \mathbf{EV}(\textit{try-B if can-try-B})$.

In other words, comparing actions in terms of their expected-utilities is equivalent to comparing conditional policies of the form *try-A if can-try-A* in terms of their expected-values. This, I take it, is the explanation for why examples led us to this definition of expected-utility. Due to the failure of action omnipotence, choosing an action is the same thing as deciding to try to perform the action if you can try to perform it. So choosing an action amounts to adopting this conditional policy, and the policy can be evaluated by computing its expected-value (or marginal expected-value). This is the intuitive rationale for the definition of expected-utility.

6. Groups of Actions

Given the reformulation of classical decision theory in terms of competition, can we take weak competition as our competition relation? In Savage's (1954) "small worlds", in which different decisions can be made wholly independently of one another, this may be the right way to understand competition. But I will argue that in general competition is more complex, for at least two reasons. The first reason turns upon the observation that in the real world we typically have a number of different decisions to make at more or less the same time. I may be deciding whether to go to the bank before lunch or after lunch, and also deciding where to go to lunch. This mundane observation creates a problem for classical decision theory because classical decision theory evaluates actions one at a time and has us choose them individually on the basis of their being optimal. The problem is that decisions can interact. Carrying out one decision may alter the probabilities and utilities involved in another decision thereby changing what action is optimal. It could be that, prior to deciding where to go to lunch, because I am very hungry the optimal decision would be to postpone going to the bank until after lunch. But if I decide to have lunch at a restaurant far from the bank and I have other things to do in that part of town that could occupy me for the rest of the afternoon, this may make it better to go to the bank before lunch. Alternatively, because I am very hungry and want to eat before going to the bank, it might be better to choose a different restaurant. The point is that actions can interfere with one another, with the result that if several actions are to be chosen, their being individually optimal does not guarantee that the group of them will be optimal. This strongly suggests that the object of decision-theoretic evaluation should be the entire group of actions rather than the individual actions.

This same conclusion can be defended in another way. Often, the best way to achieve a goal is to perform several actions that achieve it "cooperatively". Performing the actions in isolation may achieve little of value. In this case we must choose actions in groups rather than individually. To illustrate, suppose the agent's objective is to transport a ton of silver and a ton of gold from one location to another. The agent has a one-ton truck. It could fit both the gold and the silver into the truck at the same time and transport them on a single trip, but in doing so it would risk damaging the truck springs. The actions being considered are to transport the gold on a single trip, to transport the silver on a single trip, and to transport both on a single trip. We can imagine the probabilities and utilities to be such that the action with the highest expected value is that of transporting both on a single trip, even though that risks damaging the springs. However, if the agent has time to make two trips, that might be the better choice. That is, it should perform *two* actions, transporting the gold on one trip and the silver on another, rather than performing any of the single actions I am considering. This illustrates again that actions cannot always be considered in isolation. Sometimes decision-theoretic choices must be between groups of actions, and the performance of a single action becomes rational only because it is part of a group of actions whose choice is dictated by practical rationality.

These two examples illustrate two different phenomena. In the first, actions interfere with each other, changing their execution costs and hence expected values from what they would be in isolation. In the second, actions collaborate to achieve goals cooperatively, thus changing the expected values by changing the probabilities of outcomes occurring. These examples might be viewed as cases in which it is unclear that actions even have well-defined expected values in isolation. To compute the expected value of an action we must take account of the context in which it occurs. If the expected values are not well-defined, then classical decision theory cannot even be applied to these decision problems. Alternatively, if we suppose that the expected values of the actions in isolation are well-defined, then what is important about these examples is that in each case we cannot choose the group of actions by choosing the individual actions in the group on the basis of their expected values. In the first example, the expected value of the group cannot be computed by summing the expected values of the actions in the group. In the second example, the members of the group would not be chosen individually on their own strength. Rather, a pairwise comparison of actions would result in the action of transporting both the gold and silver on a single trip being chosen, and that is the intuitively wrong choice. In these examples, it is the group itself that should be the object of rational choice, and the individual actions are only derivatively rational, by being contained in the rationally chosen group of actions.

Groups of actions, viewed as unified objects of rational deliberation, are *plans*. The actions in a plan may be good actions to perform only because they are part of a good plan. It appears that the only way to get decision theory to make the right prescription in the above example is to apply it to plans rather than individual actions. The reason the agent should transport the gold alone on a single trip is that doing so is part of the plan of making two trips, and that plan is better than the plan of transporting both the gold and silver on a single trip. The plan of making two trips has a higher expected value than the plan of transporting both the gold and silver on a single trip, and that is the basis upon which it is chosen.

The reaction of most planning theorists to these examples will probably be, “Of course. So what?” But this is because I am preaching to the converted. The important point is that classical decision-theoretic mandates that choices are to be made between individual actions. I have been arguing, to the contrary, that rational choices must often be made between plan. This, I take it, is the justification for the entire enterprise of decision-theoretic planning. It does not proceed by deriving decision-theoretic planning from classical decision theory, but rather by arguing that classical decision theory is wrong in ways that require its replacement by decision-theoretic planning.

Given that rational choice requires decision-theoretic planning, how do we do it? The obvious proposal is to simply replace actions by plans in classical decision theory. *Simple plan-based decision theory* would propose that we choose between competing plans in terms of their expected-utilities. Savage (1954) seems to suggest that plans can be chosen in this way. And all of the work on decision-theoretic planning cited above takes this same form.

7. The Expected-Utility of a Plan

The failure of action omnipotence led us to define expected-utility and evaluate actions in terms of their expected-utilities. It is even more obvious that we cannot be guaranteed of being able to execute a plan, or even of being able to try to execute it (the latter because we may be unable to try to perform some of the actions it prescribes). The same considerations that led us to define the expected-utility of an action as we did lead to an analogous definition for plans. Where *p* is a plan, let us define:

expected-utility(*p*) = **MEV**(*trying to execute p if the agent can try to execute p*)

= **PROB**(*the agent can try to execute p*)·**MEV**(*try to execute p* / *the agent can try to execute p*).

What is it to try to execute a plan? This depends upon the logical structure of the plan. The simplest plans are *linear plans*, which are finite sequences of temporally ordered actions. In this section I will focus on linear plans, and in section ten I will consider plans with more complex structures. It might be supposed that trying to execute a linear plan consists of trying to execute each action it prescribes, in the right order. Somewhat surprisingly, that doesn’t work. Suppose my plan for making a cup of tea is the following: (1) heat water, (2) retrieve a tea bag, (3) place the tea bag in a teapot, (4) pour boiling water into the tea pot, (5) let the concoction sit for several minutes, (6) pour the tea from the teapot into a cup. In trying to make a cup of tea in accordance with this plan, I may start by putting the water on to boil. Then I open the cupboard to retrieve a tea bag, but discover that I am out of tea. At that point there is nothing I can do to *continue* trying to make a cup of tea. On the other hand, I did try. So trying to execute the plan does not require trying to execute each action it prescribes.

On the other hand, trying to execute a linear plan does seem to require trying to perform the first step. If I cannot even try to heat water, then I cannot try to make a cup of tea. What about the second step? Trying to execute the plan does not require that I try to perform the second step, because I may be unable to try. But suppose I can try to perform the second step. Then does trying to execute the plan require that once I have tried to perform the first step I will go on to try to perform the second step? This depends upon the plan. In the case of the plan for making tea, it is plausible to suppose that if the agent tries to heat water but fails for some reason, then he should not go on to try to perform the next step. In other words, trying to perform the second step only becomes appropriate when the agent knows that the attempt to perform the first step was successful. Such a plan can be regarded as prescribing the second step only on the condition that the first step has been successfully performed. Such plans can be constructed using “contingencies”, as described in section ten. However, not all plans will impose such conditions on the execution of their steps. If the first step is intended to achieve a result that cannot be verified by the agent at the time the plan is being executed, then the performance of the second step cannot depend upon knowing that the first step achieved its purpose. For example, the first step might consist of asking a colleague to do something later in the day. If the agent cannot verify that the colleague did as asked, the performance of the next step cannot be dependent on knowing that.

Let us define *simple linear plans* to be linear plans that impose no conditions on their steps, i.e., contain no contingencies. Simple linear plans are executed “blindly”. Once the agent has tried to perform the first step, he will try to perform the second step if he can try, and once he has tried to perform the second step he will try to perform the third step if he can try, and so on. In effect, simple linear plans are defined by the following analysis of trying to execute a simple linear plan $\langle A_1,...,A_n\rangle$:

> An agent tries to execute $\langle A_1,...,A_n\rangle$ iff:
> (1) the agent tries to perform A_1; and
> (2) for each i such that $1 < i \leq n$, if the agent has tried to perform each A_j for $j < i$, he will subsequently try to perform A_i if he can try.

So when an agent tries to execute a simple linear plan, he will try to perform the first step. Then for each subsequent step, once he has tried to perform all the earlier steps, he will try to perform the next step if he can try. If he cannot try to perform a step, the attempt to execute the plan terminates.

It follows from this that an agent *can* try to execute $\langle A_1,...,A_n\rangle$ iff he *can* try to perform A_1. To see this, consider a two-step plan $\langle A_1,A_2\rangle$. The preceding analysis yields:

> An agent tries to execute $\langle A_1,A_2\rangle$ iff:
> (1) the agent tries to perform A_1; and
> (2) if the agent has tried to perform A_1, he will subsequently try to perform A_2 if he can try.

It follows that:

> An agent can try to execute $\langle A_1,A_2\rangle$ iff:
> (1) the agent can try to perform A_1; and
> (2) if the agent has tried to perform A_1, he will subsequently be able to try to perform A_2 if he can try to perform A_2.

However, the second clause is a tautology, so it follows that the agent can try to execute $\langle A_1,A_2\rangle$ iff he can try to perform A_1. Analogous reasoning establishes the same result for *n*-step plans:

> An agent can try to execute a simple linear plan $\langle A_1,...,A_n\rangle$ iff the agent can try to perform A_1.

The most important consequence of this analysis of trying to execute a simple linear plan will be that it enables us to compute the expected-utility of the plan in terms of the conditional expected-utilities of the actions in the plan. We have seen that forming the intention to perform an action is like adopting the conditional policy of trying to perform the action if one can try, and evaluating the action in terms of its expected-utility is the same thing as evaluating the conditional policy in terms of its expected-value. Simple linear plans can also be viewed as conditional policies, although of a more complex sort. The next section will investigate "linear policies". It will be shown that the preceding analysis of trying to execute a simple linear plan has the consequence that the expected-utility of a linear plan is equal to the marginal expected-value of a corresponding "conditional linear policy", and that in turn leads to a characterization of the expected-utility of the plan in terms of the expected-utilities of its steps.

8. Linear Policies

Thus far we have considered how to define the causal probability of an outcome given a single action. In decision-theoretic planning we will often want to consider what is apt to happen if we perform several actions sequentially. Let *linear policies* be sequences of actions in which each action postdates its predecessor. First consider a linear policy consisting of two actions A_1,A_2. Computing the causal probability of an outcome is complicated by the fact that some constituents of the background of the second action may postdate the first action, and the first action can affect the probabilities of those constituents. So let B be the set of backgrounds for A_1 conjoined with those parts of the backgrounds of A_2 that predate A_1, and let B* consist of the remainders of the backgrounds for A_2. The members of B* postdate A_1. If we conceptualize the world as unfolding temporally as in figure 3 and assume that O postdates A_2, then the probability associated with a scenario should be

$$\textbf{PROB}(B_i)\cdot\textbf{PROB}(B^*_j / A_1\&B_i)\cdot\textbf{PROB}(O / A_1\&A_2\&B_i\&B^*_j).$$

Then we can define:

$$\textbf{C-PROB}_{A_1,A_2}(O) = \sum_{B\in\mathrm{B}} \textbf{PROB}(B)\cdot\sum_{B^*\in\mathrm{B}^*} \textbf{PROB}(B^* / A_1\&B)\cdot\textbf{PROB}(O / A_1\&A_2\&B\&B^*)$$

$$= \sum_{B\in\mathrm{B}} \textbf{PROB}(B)\cdot\textbf{C-PROB}_{A_2}(O / A_1\&B).$$

This definition can be generalized recursively to arbitrary sequences (for $k > 1$) of actions postdated by O:

$$\textbf{C-PROB}_{A_1,\ldots,A_k}(O) = \sum_{B\in\mathrm{B}} \textbf{PROB}(B)\cdot\textbf{C-PROB}_{A_2,\ldots,A_k}(O / A_1\&B).$$

$$\textbf{C-PROB}_{A_1,\ldots,A_k}(O / Q) = \sum_{B\in\mathrm{B}} \textbf{PROB}(B / Q)\cdot\textbf{C-PROB}_{A_2,\ldots,A_k}(O / A_1\&B\&Q).$$

Again, this calculation is what we get from propogating probabilities through scenarios in temporal order.

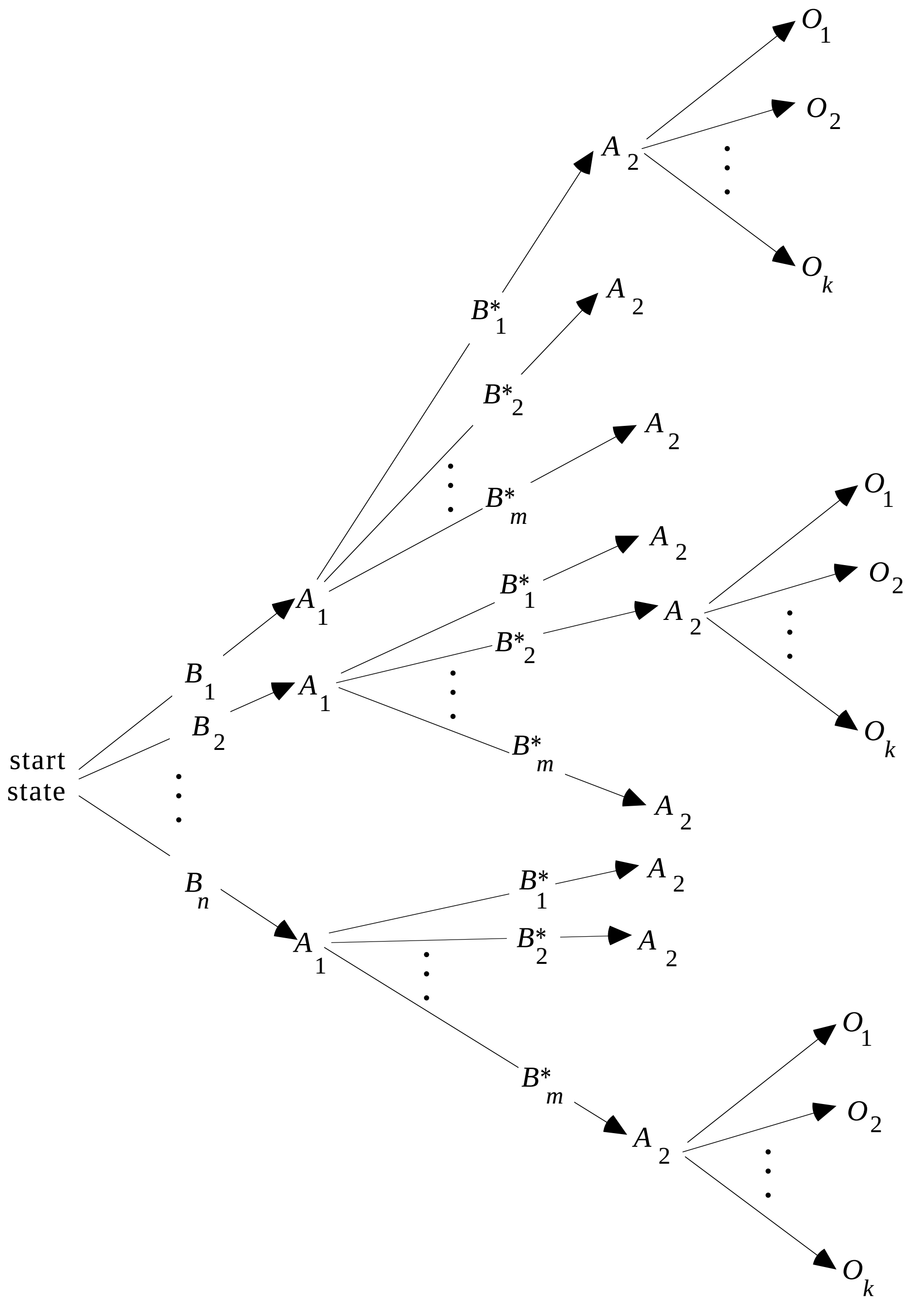

Figure 3. Scenarios with two actions

As thus-far construed, linear policies are simple sequences of actions. In decision-theoretic planning, the plans have a more complex structure. They can be viewed as *conditional linear* *policies*, which are sequences of conditional policies rather than sequences of actions. Let A_1 *if* C_1, ...,A_k *if* C_k be the policy *do* A_1 *if* C_1, *then do* A_2 *if* C_2, *then* ... . As for simple conditional policies, the probabilities of the B_i's and the probabilities of the outcomes must be made conditional on the C_i's. If C_i is false, the rest of the policy will still be executed. So on analogy to simple conditional policies, for $k > 1$ the causal probability can be defined recursively as:

$$\textbf{C-PROB}_{A_1\,if\,C_1,\ldots,A_k\,if\,C_k}(O)$$
$$= \textbf{PROB}(C_1)\cdot\sum_{B\in\mathrm{B}} \textbf{PROB}(B/C_1)\cdot\textbf{C-PROB}_{A_2\,if\,C_2,\ldots,A_k\,if\,C_k}(O/A_1\&B\&C_1)$$
$$+ \textbf{PROB}(\sim C_1)\cdot\textbf{C-PROB}_{A_2\,if\,C_2,\ldots,A_k\,if\,C_k}(O/\sim C_1).$$

$$\textbf{C-PROB}_{A_1\,if\,C_1,\ldots,A_k\,if\,C_k}(O/Q)$$
$$= \textbf{PROB}(C_1/Q)\cdot\sum_{B\in\mathrm{B}} \textbf{PROB}(B/C_1\&Q)\cdot\textbf{C-PROB}_{A_2\,if\,C_2,\ldots,A_k\,if\,C_k}(O/A_1\&B\&C_1\&Q)$$
$$+ \textbf{PROB}(\sim C_1/Q)\cdot\textbf{C-PROB}_{A_2\,if\,C_2,\ldots,A_k\,if\,C_k}(O/\sim C_1\&Q).$$

We can then define the expected-value of a conditional linear policy in terms of these causal probabilities:

$$\textbf{EV}(A_1\,if\,C_1,\ldots,A_k\,if\,C_k) = \sum_{O\in\mathrm{O}} \textbf{U}(O)\cdot\textbf{C-PROB}_{A_1\,if\,C_1,\ldots,A_k\,if\,C_k}(O).$$

Now let us apply this to simple linear plans. The expected-utility of a simple linear plan $\langle A_1,\ldots,A_n\rangle$ is **MEV**(*the agent tries to execute* $\langle A_1,\ldots,A_n\rangle$ *if he can try to execute* $\langle A_1,\ldots,A_n\rangle$). We have seen that the agent can try to execute $\langle A_1,\ldots,A_n\rangle$ iff he can try to execute A_1. It was argued in section eight that:

> An agent tries to execute $\langle A_1,\ldots,A_n\rangle$ iff:
> (1) the agent tries to perform A_1; and
> (2) for each i such that $1 < i \leq n$, if the agent has tried to perform each A_j for $j < i$, he will subsequently try to perform A_i if he can try.

It follows that:

> An agent tries to execute $\langle A_1,\ldots,A_n\rangle$ if he can try to perform A_1 iff:
> (1) the agent tries to perform A_1 if he can try to perform A_1; and

(2) for each i such that $1 < i \leq n$, if the agent has tried to perform each A_j for $j < i$, he will subsequently try to perform A_i if he can try.

Equivalently:

An agent tries to execute $\langle A_1,...,A_n\rangle$ if he can try to execute A_1 iff for each i such that $1 \leq i \leq n$, if the agent has tried to perform each A_j for $j < i$, he will subsequently try to perform A_i if he can try.

Thus adopting the plan is equivalent to adopting the conditional linear policy *try*-A_1 *if can-try*-A_1, *try*-A_2 *if can-try*-A_2 & *try*-A_1, ... , *try*-A_n *if can-try*-A_n & *try*-A_1 & ... & *try*-A_{n-1}. This is made precise by the following theorem:

Theorem 10: If $\langle A_1,...,A_n\rangle$ is a simple linear plan then

expected-utility($\langle A_1,...,A_n\rangle$)

= **MEV**(*try*-A_1 *if can-try*-A_1, ... , *try*-A_n *if can-try*-A_n & *try*-A_1 & ... & *try*-A_{n-1}).

We have seen that, for decision-theoretic purposes, deciding to perform an action amounts to deciding to try to perform it provided one can try to perform it. Because of the failure of action omnipotence, that is the most we can be sure of doing. Theorem 10 tells us that we can think of simple linear plans analogously. That is, adopting the plan amounts to deciding to try to perform its constituent actions at the appropriate times provided we can try to do so.

The point of representing plans as conditional linear policies is that we have a precise definition of the expected-value of such a policy. We can appeal to that definition to prove theorems about the expected-utility of a plan. For this purpose, let us define a different kind of conditional marginal expected-value:

MEV(A_i *if* C_i // A_1 *if* C_1,...,A_{i-1} *if* C_{i-1})
= **expected-value**(A_1 *if* C_1,...,A_i *if* C_i) – **expected-value**(A_1 *if* C_1,...,A_{i-1} *if* C_{i-1}).

Then the following fairly trivial theorem tells us that we can compute **MEV**(A_1 *if* C_1,...,A_k *if* C_k) by summing these conditional marginal expected-values:

Theorem 11: $\mathbf{MEV}(A_1 \text{ } if \text{ } C_1,...,A_k \text{ } if \text{ } C_k) = \sum_{1 \le i \le k} \mathbf{MEV}(A_i \text{ } if \text{ } C_i \,/\!/\, A_1 \text{ } if \text{ } C_1,...,A_{i-1} \text{ } if \text{ } C_{i-1})$.

If we can compute the conditional marginal expected-values $\mathbf{MEV}(A_i \text{ } if \text{ } C_i \,/\!/\, A_1 \text{ } if \text{ } C_1,...,A_{i-1} \text{ } if \text{ } C_{i-1})$, this theorem makes it easy to compute the marginal expected-value of the policy. For plans in general, these conditional marginal expected-values are usually hard to compute. However, linear plans have a particularly simple structure that often facilitates the computation. A decision-problem is *classical* iff the backgrounds are statistically independent of the actions. In that case we have a particularly simple representation of the expected-utility of a simple linear plan. First, the following theorem is proven in the appendix:

Theorem 12: If the problem is classical:

$$\mathbf{MEV}(try\text{-}A_{n+1} \text{ } if \text{ } (can\text{-}try\text{-}A_{i+1} \,\&\, the\ agent\ has\ tried\ A_1,...,A_n)$$

$$/\!/\, try\text{-}A_1 \text{ } if \text{ } can\text{-}try\text{-}A_1 ...,try\text{-}A_n \text{ } if \text{ } (can\text{-}try\text{-}A_i \,\&\, the\ agent\ has\ tried\ A_1,...,A_{n-1}))$$

$$= \mathbf{PROB}(can\text{-}try\text{-}A_1)\cdot\mathbf{PROB}(can\text{-}try\text{-}A_2 \,/\, try\text{-}A_1)\cdot ...$$

$$\cdot\mathbf{PROB}(can\text{-}try\text{-}A_{n+1} \,/\, try\text{-}A_1 \,\&\, try\text{-}A_2 \,\&...\&\, try\text{-}A_n)$$

$$\cdot\mathbf{MEV}(try\text{-}A_{n+1} \,/\, try\text{-}A_1 \&...\& try\text{-}A_n \& can\text{-}try\text{-}A_{n+1}).$$

If we define:

$$\mathbf{expected\text{-}utility}(A_{n+1} \,/\, A_1,...,A_n)$$

$$= \mathbf{PROB}(can\text{-}try\text{-}A_1)\cdot\mathbf{PROB}(can\text{-}try\text{-}A_2 \,/\, try\text{-}A_1)\cdot ...$$

$$\cdot\mathbf{PROB}(can\text{-}try\text{-}A_{n+1} \,/\, try\text{-}A_1 \,\&\, try\text{-}A_2 \,\&...\&\, try\text{-}A_n)$$

$$\cdot\mathbf{MEV}(try\text{-}A_{n+1} \,/\, try\text{-}A_1 \&...\& try\text{-}A_n \& can\text{-}try\text{-}A_{n+1})$$

it follows from theorems 11 and 12 that:

Theorem 13: If the problem is classical:

$$\textbf{expected-utility}(\langle A_1,\ldots,A_n\rangle) = \sum_{1 \le i \le n} \textbf{expected-utility}(A_i / A_1,\ldots,A_{i-1}).$$

In theorem 13, $\textbf{expected-utility}(A_i / A_1,\ldots,A_{i-1})$ can be regarded as the amount of utility trying to perform A_i (if one can) can be expected to add after the preceding steps of the plan have been attempted.

9. Conditional and Nonlinear Plans

Section eight focused on "simple linear" plans, where if the agent is unable to try to perform a step then he stops trying to execute the rest of the plan, but otherwise he continues. However, most plans have more complex structures. To begin with, plans need not be linear. Some steps of the plan can be unordered with respect to others. Using an example from Russell and Norvig (1995), a plan to put on your shoes and socks might look as in figure 4. Here *put on right sock* is ordered before *put on right shoe*, but it is unordered with respect to either *put on left sock* or *put on left shoe*. When the plan is actually executed, the steps will be performed in some order, but the decision of which steps to perform first can be left until the time of execution. Such plans are *nonlinear plans*.

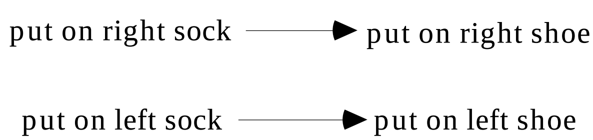

Figure 4. A nonlinear plan

A *linearization* of a nonlinear plan is a linear plan that results from adding ordering constraints to the nonlinear plan to linearly order the plan steps. The expected-utility of a nonlinear plan ought to be determined by the expected-utilities of its linearizations. If a planning agent discovers that some linearizations have lower expected-utilities than others, then he should add ordering constraints that preclude those linearizations. So it is reasonable to define the expected-utility of a nonlinear plan to be the minimum of the expected-utilities of its linearizations. Ideally, all the linearizations should have the same expected-utility.

Nonlinear plans are generally constructed by merging smaller linear plans into a single plan. For example, the above plan might be produced by merging a plan to put on my right shoe and sock and a plan to put on my left shoe and sock. When plans are merged, the result will normally be a nonlinear plan because steps drawn from different constituent subplans will usually be unordered with respect to each other.

The linearizations of a nonlinear plan will not normally be simple linear plans. If we are unable to try to perform a step of one of the constituent plans out of which the nonlinear plan is constructed, that should abort execution of the rest of that constituent plan, but may not abort execution of the other constituent plans. For example, if I cannot find my right sock, I may still try to don my left sock. What this illustrates is that in general, plans will incorporate a dependency relation on steps, and if the agent is unable to try to perform a step, he will only try to execute the remaining steps that do not depend upon the failed step. A step can have several purposes in a plan, and for each purpose it may depend upon a different set of steps. We can handle this by talking about *dependency-sets* for plan-steps. The *initial steps* of a plan will have empty dependency-sets and no steps will be ordered before them. Let us define recursively that a step is *cancelled* iff either the agent was unable to perform it or every dependency-set for it contains some some cancelled step. Then we might try defining:

> An agent tries to execute a plan iff each step *S* of the plan, if the agent has tried to perform all of the steps in some dependency-set for *S* and all the uncancelled steps ordered before *S* in the plan, then he will subsequently try to perform *S* if he can try.

This has the consequence that in order to try to execute a plan, the agent must try to execute all of the initial steps. However, this definition does not quite capture what we want. For example, if we apply this to simple linear plans, each step S_{i+1} will have a single dependency-set containing only its immediate predecessor S_i. The preceding analysis of trying to execute a plan would then make it equivalent to executing a linear policy whose clauses have the form *try*-A_n *if can-try*-A_n & *try*-A_{n-1}. However, as we have seen, the clauses of the linear policy ought to have the form *try*-A_n *if can-try*-A_n & *try*-A_1 & ... & *try*-A_{n-1}. This will make an important difference to the policy. The policy whose clauses have the form *try*-A_n *if can-try*-A_n & *try*-A_{n-1} would have us consider the possibility that although the agent was unable to try to perform some of the earlier steps, and so the later steps were not "called" by the plan, he nevertheless performs *try*-A_{n-1} for some other reason. The policy would then require him to go on and perform *try*-A_n if he can. This is not required by the policy whose clauses have the form *try*-A_n *if can-try*-A_n & *try*-A_1 & ... & *try*-A_{n-1}. Only the latter provides the correct analysis of what it is to try to execute the simple linear policy.

We can express the preceding observations by saying that the policy should only require the agent to try to perform *try*-A_n if *can-try*-A_n & *try*-A_{n-1} and *try*-A_{n-1} was "called" by the policy. In the simple linear policy, *try*-A_{n-1} is called just in case the agent tried to perform all the previous steps, but in nonlinear plans the calling relation is more complex. It can be defined recursively as follows:

> A step *S* is *called* iff for some dependency-set, (1) every step in the dependency-set is called and the agent tried to perform it, and (2) for every step ordered before *S* in the plan, either it is cancelled or it has been called and the agent tried to perform it if he could try.

Then we can repair the flawed analysis of trying to execute a plan as follows:

> An agent tries to execute a plan iff for each step *S* of the plan, if *S* is called then the agent will try to perform *S* if he can.

This has the consequence that in order to try to execute the plan, the agent must try to perform all of the initial steps.

To produce plans with reasonable expected-utilities, it will be important to add a further complication to the logical structure of plans. In planning what to do, we often lack knowledge that would help us make a better decision. For example, in planning a route for driving from one point to another, I may be unsure whether a certain road is under construction. Ideally, I should plan to go one way if the road is under construction, but a different way if it is not. We handle this by planning for both possibilities, and then decide at the time we execute the plan which way to go. This can be accommodated by inserting *contingencies* into a plan.[2] A contingency is labeled with an epistemic condition — normally that the agent does or does not believe some proposition *P* — and then some of the subsequent plan steps are made dependent on it. At execution time, those plan steps are only executed if the agent satisfies the epistemic condition. Let us say that the agent *satisfies* a contingency if he satisfies the condition labeling it.

To accommodate the dependency of plan-steps on contingencies, we can include contingencies in the dependency-sets. We expand the definition of cancellation by saying that a step is *cancelled* iff either the agent was unable to perform it or every dependency-set for it contains some cancelled step or some contingency the agent does not satisfy. Then we can revise the definition of a step being called as follows:

---

[2] This basic idea derives originally from work on classical contingency planning by Warren (1976), Peot and Smith (1992), and Pryor and Collins (1996). Attempts to combine contingency planning with decision-theoretic planning have been made by Draper, Hanks, and Weld (1994), Blythe and Veloso (1997), and Onder and Pollack (1997, 1999). For a survey of the ways in which decision-theoretic plans can be improved by adding contingencies, see my (2001a).

A step *S* is *called* iff (1) for some dependency-set, all the contingencies are satisfied, and every step in the dependency-set is called and the agent tried to perform it, and (2) for every step ordered before *S* in the plan, either it is cancelled or it has been called and the agent tried to perform it if he could try.

The preceding definition has the consequence that initial steps may not be called, because they may depend upon contingencies that are not satisfied. If none of them are called, then even if the agent tries to perform every step of the plan that is called, he may do nothing. If he does nothing, he has not tried to execute the plan. So we must modify the definition of trying to execute a plan as follows:

An agent tries to execute a plan iff (1) some initial step is called, and (2) for each step *S* of the plan, if *S* is called then the agent will try to perform *S* if he can.

Corresponding to each step *S* of a (possibly nonlinear) plan is the conditional policy *try to perform S if (1) you can try, (2) you satisfy all the contingencies in some dependency-set for S, (3) all the steps in the dependency-set have been called and you have already tried to perform them, and (4) all the uncancelled steps ordered before S in the plan have been called and you have tried to perform them.* However, collecting these together does not generate a conditional linear policy because the steps need not be linearly ordered. If a plan is nonlinear then there will be points in the course of its execution at which the agent has a choice of what to do next. These will be cases in which the conditions of more than one of these conditional policies are satisfied simultaneously. Accordingly, we cannot apply results pertaining to conditional linear policies directly to the evaluation of the plan.

This problem can be circumvented by looking at the linearizations of the plan. The plan then corresponds to the conditional linear policy that is the sequence of the conditional policies corresponding to the steps of the plan, and the expected-utility of the linearization is the marginal expected-value of this policy. This enables us to use general theorems about conditional linear policies to characterize the expected-utility of a plan.

Recall that causal probability was defined for conditional linear policies as follows:

$$\textbf{C-PROB}_{A_1\ if\ C_1,\ldots,A_k\ if\ C_k}(O)$$
$$= \textbf{PROB}(C_1)\cdot\sum_{B\in\mathbf{B}} \textbf{PROB}(B/C_1)\cdot\textbf{C-PROB}_{A_2\ if\ C_2,\ldots,A_k\ if\ C_k}(O/A_1\&B\&C_1)$$
$$+\ \textbf{PROB}(\sim C_1)\cdot\textbf{C-PROB}_{A_2\ if\ C_2,\ldots,A_k\ if\ C_k}(O/\sim C_1).$$

$$\textbf{C-PROB}_{A_1\ if\ C_1,\ldots,A_k\ if\ C_k}(O/Q)$$

$$= \textbf{PROB}(C_1/Q)\cdot\sum_{B\in\mathrm{B}} \textbf{PROB}(B/C_1\&Q)\cdot\textbf{C-PROB}_{A_2\ if\ C_2,\ldots,A_k\ if\ C_k}(O/A_1\&B\&C_1\&Q)$$

$$+ \textbf{PROB}(\sim C_1/Q)\cdot\textbf{C-PROB}_{A_2\ if\ C_2,\ldots,A_k\ if\ C_k}(O/\sim C_1\&Q).$$

For simple linear policies, scenarios as diagrammed in figure 3 have the form

$$B_1 \rightarrow A_1 \rightarrow B_2 \rightarrow A_2 \rightarrow \ldots \rightarrow B_k \rightarrow A_k.$$

Scenarios for conditional linear policies are more complex, because they must take account of whether the conditions are satisfied. Each condition may be either satisfied or unsatisfied, and if it is unsatisfied then the corresponding action is not included in the scenario. This means that a scenario for a conditional linear policy looks like this:

$$(\sim)C_1 \rightarrow B_1 \rightarrow (A_1) \rightarrow (\sim)C_2 \rightarrow B_2 \rightarrow (A_2) \rightarrow \ldots \rightarrow (\sim)C_k \rightarrow B_k \rightarrow (A_k)$$

where each tilde can be present or absent, and A_i is included in the scenario iff the tilde is absent on C_i. A scenario is characterized by the set of unnegated C_i's. For example, the following is a scenario:

$$C_1 \rightarrow B_1 \rightarrow A_1 \rightarrow \sim C_2 \rightarrow B_2 \rightarrow C_3 \rightarrow B_3 \rightarrow A_3$$

Given a scenario S, let C_S be the conjunction of the C_i's, $\sim C_i$'s, and B_i's in the scenario, and let A_S be the conjunction of the actions in the scenario. Defining the probability of a scenario as $\textbf{C-PROB}_{A_1\ if\ C_1,\ldots,A_k\ if\ C_k}(C_S)$, we get:

Theorem 14: If SC is the set of scenarios for the policy $A_1\ if\ C_1,\ldots,A_k\ if\ C_k$ relative to O then

$$\textbf{C-PROB}_{A_1\ if\ C_1,\ldots,A_k\ if\ C_k}(O/Q)$$

$$= \sum_{S\in\mathrm{SC}} \textbf{C-PROB}_{A_1\ if\ C_1,\ldots,A_k\ if\ C_k}(S/Q)\cdot\textbf{PROB}(O/C_S\&A_S\&Q).$$

It then follows that:

Theorem 15: If SC is the set of scenarios for the policy $A_1\ if\ C_1,\ldots,A_k\ if\ C_k$ then:

$$\mathbf{MEV}(A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k) = \sum_{S \in \mathrm{SC}} \textbf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(S)\cdot\mathbf{MEV}(S).$$

Thus the marginal expected-value of a conditional linear policy is a weighted average of the marginal expected-values of its scenarios.

The next theorem tells us that the marginal expected-value of a scenario is the sum of the marginal expected-values of its actions in the context of the scenario:

Theorem 16: If S is a scenario and $A_1,\ldots,A_n$ are the actions it prescribes listed in temporal order then:

$$\mathbf{MEV}(S) = \sum_{1\leq i \leq n}\mathbf{MEV}(A_i / A_1 \,\&\ldots\&\, A_{i-1} \,\&\, C_S).$$

Theorems 16 and 17 together tell us how to compute the marginal expected-value of a linear policy (and hence how to compute the expected-utility of a decision-theoretic plan) in terms of the marginal expected-values of the actions constituting the policy in the possible scenarios of the policy. Unfortunately, the computation this prescribes will often be very difficult. It requires us to compute marginal expected-values for every scenario separately, and compute the marginal expected-value of the policy or plan as a weighted average of the marginal expected-values of the scenarios. There can be a very large number of scenarios. Thus efficient decision-theoretic planning requires more efficient ways of computing expected-utilities. However, I will not pursue that topic here.

10. Choosing Between Plans

We have seen that rational choices must often be between plans rather than individual actions. “Simple plan-based decision theory” proposes that we can characterize rational choice by reconstruing classical decision theory to apply to plans rather than actions. In pursuing this suggestion, we have seen how to define the expected-utility of a plan, and the proposal is now that a rational agent can choose between plans on the basis of their expected-utilities. However, just as for actions, we need not choose between plans unless they are in some sense in competition. If two plans are not in competition, we can simply adopt both. So to construct a plan-based theory of rational choice, we need an account of when plans compete in such a way that a rational choice should be made between them.

Competing plans should be plans that we must choose between, rather than adopting both.

A sufficient condition for this is that executing one of the plans makes it impossible to execute the other one, i.e., the plans *compete strongly*. We want alternatives to be actions we should, rationally, choose between. That is, we should choose one but not more than one. This can be the result of much weaker relations than strong competition. For example, consider an office robot instructed to copy a document and deliver the copy to one of two offices. It has to choose between the plan to copy the document and deliver it to office-1 and the plan to copy the document and deliver it to office-2. It *could* do both. The plans do not compete strongly. But nevertheless, they are in competition. We do not want the robot to adopt both plans because a copy of the document only needs to be delivered to one office. We might capture this with a notion of weak competition — let us say that two plans *compete weakly* iff the plan that results from merging the two plans into a single plan has a lower expected value than at least one of the original plans. Delivering copies of the document to both offices has a lower expected value than delivering a copy to just one because (1) the execution cost is essentially the sum of the execution costs of the two separate plans, but (2) only one delivery is required, so the payoff from delivering copies to both offices is no greater than the payoff from delivering a copy to just one. It might be proposed, then, that two plans are competitors iff they compete weakly. An appeal to weak competition does not generate a theory with quite the same structure as classical decision theory. The problem is that classical decision theory assumes we have a set of alternative actions, and prescribes choosing an optimal member of the set. However, weak competition doesn't generate a set of alternatives that are pairwise competitors. This is because weak competition is not transitive. $Plan_1$ may compete weakly with $plan_2$, and $plan_2$ with $plan_3$, without $plan_1$ competing weakly with $plan_3$. Thus if we simply pick an plan and let the set of alternatives be the set of all plans competing weakly with the given plan, it does not follow that other members of the set of alternatives will be in competition. It may be desirable to perform several of those other plans together. For instance, suppose I am planning a wedding and want to select something borrowed and something blue, but it is undesirable to select two borrowed things or two blue things. If *x* and *y* are borrowed, and *y* and *z* are blue, then selecting *x* competes weakly with selecting *y*, and selecting *y* competes weakly with selecting *z*, but selecting *x* does not compete weakly with selecting *z*. Accordingly, it seems that what classical decision theory really ought to say is:

> It is rational to adopt (decide to execute) a plan iff it has no weak competitor with a higher expected-utility.

To evaluate this proposal, let us first observe that a cognitively sophisticated autonomous agent operating in a complex environment is not faced with a single fixed planning problem. First, its beliefs will change as it acquires experience of its environment and as it has time for further reasoning. This will affect what solutions are available for its planning problems. Second, as it acquires more knowledge of its environment, its goals may change (see section fifteen). We cannot expect the agent to redo all of its previous planning each time it acquires new knowledge or new goals, so planning must produce lots of *local plans*. These are small plans of limited scope aiming at disparate goals.

This gives rise to a very fundamental problem for plan-based decision theory. The problem is that simple plan-based decision theory cannot be applied to local plans that are generated in this way. The theory is formulated in terms of whether there exists (in logical space) a competing plan with a higher expected value. The problem is that there will almost always exist such a competing plan. To illustrate, think again about the office robot that is choosing between copying the document and delivering it to office-1 and copying the document and delivering it to office-2. Suppose the former has the higher expected value (perhaps because it has a lower execution cost). This implies that the plan of delivering a copy to office-2 is not rationally adoptable, but it does not imply that the plan of delivering a copy to office-1 is adoptable, because some other plan with a higher expected value may compete with it. And we can generally construct such a competing plan by simply adding steps to the earlier competing plan. For this purpose, we select the new steps so that they constitute a subplan achieving some valuable unrelated goal. For instance, we can consider the plan of delivering a copy to office-2 and then recharging the battery. This plan still competes with the plan of delivering a copy to office-1, but it has a higher expected value. Thus the plan of delivering a copy to office-1 is not rationally adoptable. However, the competing plan is not rationally adoptable either, because it is trumped by the plan of delivering a copy to office-1 and then later recharging the battery.

It seems clear that given two competing plans P_1 and P_2, if the expected-utility of P_1 is greater than that of P_2, the comparison can generally be reversed by finding another plan P_3 that pursues unrelated goals and then merging P_2 and P_3 to form P_2+P_3. If P_3 is well chosen, this will have the result that P_2+P_3 still competes with P_1 and the expected-utility of P_2+P_3 is higher than the expected-utility of P_1. If this is always possible, then there are no optimal plans and simple plan-based decision theory implies that it is not rational to adopt any plan.

In an attempt to avoid this problem, it might be objected that P_2+P_3 is not an appropriate object of decision-theoretic choice, because it merges two unrelated plans. However, recall the second example used to motivate the application of decision theory to plans rather than actions — the example of transporting a ton of gold and a ton of silver. The plan we wanted to adopt in preference to transporting either the gold, the silver, or both on a single trip, was the plan to transport the gold on one trip and the silver on another trip. However, this plan is constructed by merging two unrelated plans for achieving unrelated goals. If we are not allowed to construct such merged plans, decision theory will not produce the intuitively correct prescription in this example.

The inescapable conclusion is that the rational adoptability of a local plan cannot require that it have a higher expected-utility than all its competitors. The problem is that plans can have rich structures and can pursue multiple goals, and as such they are indefinitely extendable. We can almost always construct competing plans with higher expected-utilities by adding subplans pursuing new goals. Thus there is no way to define optimality so that it is reasonable to expect there to be optimal plans. Hence simple plan-based decision-theory fails.

I have heard it suggested that this problem does not arise for Markov decision process planning. However, there are two ways of viewing MDP planning. We could think of the entire world as a single MDP and try to produce a single "universal" plan governing the agent's actions for all time. But that problem is completely intractable — see the next section. Alternatively, we might use MDP planning techniques to find local plans (policies) by separating out toy problems, but here the preceding difficulty recurs. You can compare the plans produced for a single toy problem in terms of their expected values, but if you expand the problem and consider more sources of value, more possible actions, etc., in effect looking at a bigger toy problem, an optimal policy for the larger problem may not contain the optimal policy for the subproblem. So the problem recurs. It arises independently of the kind of planning algorithms employed.

In constructing a plan-based decision theory, we want to define a relation of "better plan" that can drive rational decisions. It is natural to suppose that a plan should only be adopted if it has no competitor that is better than it. If we make the latter assumption and we identify "better plan" with "plan with a higher expected value", then it will normally be the case that no local plan is rationally adoptable. This result is absurd, so such an identification must be incorrect.

11. Universal Plans

There is a way of trying to save plan-based decision theory from the preceding objection. The argument that led to the conclusion that plans cannot be selected for adoption just by comparing their expected-utilities turned upon always being able to extend a plan by merging it with a subplan for achieving an additional goal. For plans as ordinarily conceived, this assumption is unproblematic. But there is one way of avoiding the argument — consider only "universal plans". These are plans prescribing what the agent should do for all the rest of its existence. Universal plans cannot be extended by adding subplans for new goals. Because universal plans include complete prescriptions for what to do for the rest of the agent's existence, any two universal plans will make different prescriptions, and so will strongly compete. It seems initially quite plausible to suppose that universal plans can be compared in terms of their expected-utilities, and that a universal plan is rationally adoptable iff it is optimal, i.e., iff no other universal plan has a higher expected-utility.

Savage (1954) toys with the idea that rational decisions should be between universal plans, but he rejects it for the obvious reason. The real world is too complex. No agent with realistic computational limitations could possibly construct a universal plan prescribing the optimal action for every possible circumstance. To illustrate the severity of this problem, consider a cognitively challenged agent that can only take account of 300 independent (two-valued) properties of situations. Obviously, the number of properties human beings take into account in the real world is orders of magnitude greater. But even for such a cognitively impoverished agent, there will be 2^{300} possible situations that it can distinguish between and must plan for. A universal plan must prescribe the optimal action for each of these 2^{300} possible situations. 2^{300} is an immense number — approximately equal to 10^{90}. To appreciate just how large this number is, it has been estimated that the number of elementary particles in the universe is 10^{78}. So this would require planning for 12 orders of magnitude more possible situations than there are elementary particles in the universe. And this is for just 300 properties. Of course, if the world is particularly well behaved, then it may be possible to give general descriptions of optimal actions in large classes of situations rather than making explicit prescriptions for each situation (i.e., factor the universal plan). But it is preposterous to suppose that even that will enable a real agent to find a universal plan for dealing with every possible situation it may encounter in the real world. Real agents will not be able to find universal plans.

So what? Theories of rationality are often taken to be theories of ideally rational agents, unconstrained by realistic computational limitations. According to this point of view, if ideally rational agents have to have infinite memory and computing power, and real agents cannot, then so much the worse for real agents — they are necessarily irrational. One can, of course, choose to employ the term “theory of rationality” in this way, but so conceived it is hard to see what a theory of rationality is good for. At the beginning of the paper, I introduced theories of rational action somewhat differently — they are theories of how a real agent should, rationally, go about deciding what actions to perform at any given time. A theory that requires an agent to do something that is impossible cannot be a correct theory of rationality. Real agents cannot compute universal plans, so the theory of rationality cannot require them to.

12. The Real World vs. Toy Problems

My main objection to classical decision theory is that actions cannot be chosen in isolation. My main objection to simple plan-based decision theory is that optimality must always be relative to a set of alternatives, and there is no way of selecting an appropriate set of decision-theoretic alternatives in such a way that there will always or even normally be an optimal choice. A secondary objection is that even in those rare cases where there might be an optimal plan, real agents cannot be expected to find them. But suppose that the choice of which plan to perform could always made from a manageably small (finite) set of alternatives. Then optimality would be well-defined, and the prescription to choose an optimal plan would seem sensible. It would seem reasonable to require the agent to evaluate each of the alternative plans and choose an optimal one. However, outside of toy problems, this supposition is indefensible. Plans are mathematical or logical constructions of unbounded complexity. In the real world, if we can construct one plan for achieving a goal, we can typically construct infinitely many. In designing autonomous agents, talk of "planning domains" is inappropriate. The only planning domain is the whole world.

It should be emphasized that this is a claim about autonomous agents operating in the real world — not in artificial environments. I want to build Isaac Asimov's *I-Daneel* and Arthur C. Clarke's *HAL*. I take it that this is one of the seminal aspirations of AI, although at this time no one is in a position to actually try to do it. For some concrete applications it is possible to constrain the planner's environment sufficiently to turn the problem into what is in effect a toy problem. This may be possible if the planning problem can be fixed from inception and either (1) local plans aiming at distinct goals do not interact, or (2) it is possible to construct universal plans with manageably short time horizons and manageably few actions. In such a world, optimization will be both well-defined and desirable. But the most important point of this paper is that the techniques that work for toy problems will, for purely logical reasons, not scale up to the problem of building an autonomous agent operating in the real world. For such agents, there is no reason to expect there to be optimal plans. The difference between toy problems and the real world is not just a difference of degree. Planning must work entirely differently.

It follows that we can neither define optimality nor determine whether a plan is optimal by just comparing it with a small pre-computed set of alternative plans. We are going to have to inspect infinitely many plans in order to determine whether a given plan is optimal, and it will usually turn out that none is. This cannot have the consequence that we should not adopt any plan, so it follows that finding optimal plans cannot be the proper aim of rational choice. On the other hand, there has to be some way of making sense of one plan being better than another, and it seems this must have something to do with the probabilities and utilities of outcomes. Further, it seems that there must be some account of rational choice that proceeds in terms of this relation of one plan being better than another. In the next section, I will propose an alternative to simple plan-based decision theory.

13. When is One Plan Better than Another?

We want to define a notion of "better plan" which is of use in deciding whether a plan should be adopted. I have argued that this cannot be cashed out as the first plan merely having a higher expected value than the second. To get a grip on this notion, let us think about plan adoption in rational agents. First consider the limiting case in which an agent has no background of adopted plans, and a new plan is constructed. Should the new plan be adopted? The basic insight of classical decision theory is that what makes a course of action (a plan) good is that it will, with various probabilities, bring about various value-laden states, and the cost of doing this will be less than the value of what is achieved. This can be assessed by computing the expected value of the plan. In deciding whether to adopt the plan, all the agent can do is compare the new plan with the other options currently available to it. If this is the only plan the agent has constructed, there is only one other option — do nothing. So in this limiting case, we can evaluate the plan by simply comparing it with doing nothing. This is the same thing as asking whether its expected value is positive.

Things become more complicated when the agent has already adopted a number of other plans. This is for two reasons. First, the new plan cannot be evaluated in isolation from the previously adopted plans. Trying to execute the previous plans may affect both the probabilities and the utilities employed in computing the expected value of the new plan. For example, if the new plan calls for the agent to perform a an operation that requires large amounts of battery power, the probability of the agent being able to do that may normally be fairly high. But if other plans the agent has adopted will result in the agent having depleted its battery power, then the probability of being able to perform the operation may be lower. So the probabilities can be affected by the context provided by the agent's other plans. The same thing is true of the values of goals. Suppose the new plan is a plan for eating a sumptuous meal. In the abstract, this may have a high value, but if it is performed in a context in which the agent's other plans include participation in a pie-eating contest immediately before dinner, the value of the sumptuous meal will be seriously diminished. Execution costs can be similarly affected. If the new plan prescribes transporting an object from one location to another in a truck, this will be more costly if a previous plan moves the truck to the other side of town.

Clearly, the expected value of the new plan must be computed "in the context of the agent's other plans". But what does that mean? Roughly, the probabilities and utilities should be conditional on the situation the agent will be in as a result of having adopted and tried to execute parts of the other plans. However, there isn't just one possible situation the agent might be in, because the other plans will normally have their results only probabilistically.

The second reason it becomes more complicated to evaluate a new plan when the agent already has a background of adopted plans is that the new plan can affect the value of the old plans. If an old plan has a high probability of achieving a very valuable goal but the new plan makes the old plan unworkable, then the new plan should not be adopted. Note that this is not something that is revealed by just computing the expected value of the new plan.

We have seen that normal planning processes produce local plans. How should the agent decide whether to adopt a new local plan? The decision must take account of both the effect of previously adopted plans on the new plan, and the effect of the new plan on previously adopted plans. We can capture these complexities in a precise and intuitively appealing way by defining the concept of the agent's *master plan*. This is the result of merging all of the agent's adopted plans into a single plan. A master plan is a *global plan* for achieving all (or as many as possible) of the agent's current goals simultaneously.

The significance of the master plan is that it can be used as a tool for evaluating local plans. This turns upon the fact that, unlike local plans, master plans can be compared in terms of their expected values. The expected value of the master plan is our expectation of how much better the world will be if we adopt that as our master plan. Thus one master plan is better than another iff it has a higher expected value. Equivalently, rationality dictates that if an agent is choosing between two master plans, he should choose the one with the higher expected value.

If the only way an agent had of finding a master plan with a higher expected value than its current one was to plan all over again from scratch and produce a new master plan essentially unrelated to its present master plan, the task would be so formidable as to be computationally impossible. The only way resource-bounded agents can construct and improve upon master plans reflecting the complexity of the real world is by constructing or modifying them incrementally. When trying to improve its master plan, rather than throwing it out and starting over from scratch, what an agent must do is try to improve it piecemeal, leaving the bulk of it intact at any given time. This is where local plans enter the picture. Normal planning processes produce local plans. The significance of local plans is that they represent the building blocks for master plans. Earlier, we encountered the problem of how to evaluate a newly constructed local plan, given that we must take account both of its effect on the agent's other plans, and the effect of the agent's other plans on the new plan. We are now in a position to answer that question. The only significance of local plans is as constituents of the master plan. When a new local plan is constructed, what we want to know is whether the master plan can be improved by adding the local plan to it. Thus when a new plan is constructed, it can be evaluated in terms of its impact on the master plan. We merge it with the master plan, and see how that affects the expected value of the master plan. Let us define the *marginal expected value* of the local plan P to be the difference its addition makes to the master plan M:

$$\mathbf{MEU}(P,M) = \mathbf{expected\text{-}utility}(M+P) - \mathbf{expected\text{-}utility}(M).$$

If the marginal expected-utility is positive, adding the local plan to the master plan improves the master plan, and so in that context the local plan is a good plan. Furthermore, if we are deciding which of two local plans to add to the master plan, the better one is the one that adds more value to the master plan. So viewed as potential additions to the master plan, local plans should be evaluated in terms of their marginal expected-utilities, not in terms of their expected-utilities simpliciter.

14. Locally Global Planning

My proposal is that the "better plan" relation is a three-place relation, comparing the marginal expected-utilitys of plans relative to master plans. But how exactly do we use this relation to formulate a theory of rational choice? It appears that the aim of plan search is to construct local plans and use them to improve the master plan. It may at first occur to one that the objective should be to find an optimal master plan. But that cannot be right, for two reasons. First, it is unlikely that there will ever be optimal master plans that are smaller than universal plans. If a master plan leaves some choices undetermined, it is likely that we can improve upon it by adding decisions regarding those choices. But as we have seen, it is not possible for real agents to construct universal plans, so that cannot be required for rational choice.

The idea that rationality requires choosing optimal master plans is a holdover from classical decision theory. Classical decision theory envisages a kind of "ideal rationality" where an agent can survey all possible courses of action and choose an optimal one. But that is a computationally impossible ideal. Real rationality — the rules governing rational cognition in real agents operating in complex environments — must set different standards. Most work in AI has assumed that an agent can complete all relevant reasoning before deciding how to act. But outside of toy problems, that will never be the case. Assuming that the agent's reasoning about the world involves at least full first-order logic, and more likely some defeasible (nonmonotonic) reasoning about its environment, that reasoning will not produce a recursive set of conclusions and so will in general be non-terminating.[3] Even if the agent is only engaging in classical planning, if it has to reason about its environment to detect threats to causal links then the set of threats will not generally be recursive, and I showed in my (1998) that this makes the set of ⟨*problem,solution*⟩

[3] I showed in my (1995) that even if the set of reasoning schemas available to a defeasible reasoner are "well behaved", the set of conclusions will only be Δ_2 in the arithmetic hierarchy.

pairs not even recursively enumerable. In general, reasoning will be non-terminating. There will be no point at which an agent has exhausted all possibilities in searching for plans. Despite this, agents must take action. They cannot wait for the end of a non-terminating search before deciding what to do, so their decisions about how to act must be directed by the best plans found to date — not by the best possible plans that *could* be found. The upshot is that plan adoption must be defeasible. Agents must work with the best knowledge currently available to them, and as new knowledge becomes available they may have to change some of their earlier decisions. If we only test agent designs on toy problems, even big toy problems, we are apt to be led to architectures that cannot handle this rather fundamental observation.

This point is fairly obvious, and yet it completely changes the face of decision-theoretic planning. The objective cannot be to find optimal master plans. First, non-terminating reasoning may produce better and better master plans without limit, so there may be no optimal master plans. Second, even if there were optimal master plans the agent would have no way of knowing it has found one until all of the non-terminating reasoning is completed. Planning and plan adoption must be done defeasibly, and actions must be chosen by reference to the current state of the agent's reasoning at the time it has to act rather than by appealing to the idealized but unreachable state that would result from the agent completing all possible reasoning and planning. Agents begin by finding good plans. The good plans are "good enough" to act upon, but given more time to reason, good plans might be supplanted by better plans.[4] The agent's master plan evolves over time, getting better and better, and the rules for rationality are rules directing that evolution, not rules for finding a mythical endpoint. Accordingly, a decision-theoretic planner should implement rules for continually improving the master plan rather than implementing a search for the endpoint, i.e., a search for optimal master plans. We might put this by saying that a decision-theoretic planner should be an *evolutionary planner*, not an optimizing planner. An evolutionary planner will be implemented as an infinite loop rather than a terminating search program. A program for evolutionary planning will systematically direct an agents "entire life" rather than just discrete segments aimed at local goals.

The upshot of all this is that rational choice becomes a theory of how to construct local plans and use them to improve the global plan — the master plan. I call this *locally global planning.* As a first approximation, we might try to formulate locally global planning as follows:

---

[4] This is reminiscent of Herbert Simon's (1955) concept of "satisficing", but it is not the same. Satisficing consists of setting a threshold and accepting plans whose expected-utilities come up to the threshold. The present proposal requires instead that any plan with a positive expected-utility is defeasibly acceptable, but only defeasibly. If a better plan is discovered, it should supplant the original one. Satisficing would have us remain content with the original.

> It is rational for an agent to adopt a plan iff its marginal expected-utility is positive, i.e., iff adding it to the master plan increases the expected-utility of the master plan.

However, for two reasons, this formulation is inadequate. The first reason is that adding a new plan may only increase the expected value of the master plan if we simultaneously delete conflicting plans. For example, suppose the agent has adopted one plan for achieving a goal, and then discovers a better plan. The master plan will not usually be improved by simply adding the new plan to it. That would result in the agent's having two plans for achieving the goal. It will normally be better to adopt the new plan and simultaneously delete the earlier plan.

The second reason the preceding formulation is inadequate is that plans may have to be added in groups rather than individually. Recall again the example of transporting the gold and silver to a common destination in a truck. The plan to deliver the gold and silver on a single trip, by virtue of achieving both goals (and taking account of the possible damage to the truck), had a higher expected value than any single plan with which it competes, e.g., the plan to deliver the gold without delivering the silver. What is better than adopting the plan to deliver them both on a single trip is adopting the two separate plans to deliver the gold on one trip and deliver the silver on another trip. So suppose the agent first adopts the plan to deliver both the gold and the silver on a single trip. Then it occurs to the agent that it could make two trips. The change it should make to the master plan at that point involves deleting the plan to deliver the gold and silver on a single trip, and adding two other plans — the plan to deliver the gold on one trip and the plan to deliver the silver on another trip.

In general, a change to the master plan may consist of deleting several local plans and adding several others. Where M is a master plan and C a change, let $M\Delta C$ be the result of making the change to M. We can define the *marginal expected-utility of a change C* to be the difference it makes to the expected-utility of the master plan:

$$\mathbf{MEU}(C,M) = \mathbf{expected\text{-}utility}(M\Delta C) - \mathbf{expected\text{-}utility}(M).$$

The *principle of locally global planning* can then be formulated as follows:

> It is rational for an agent to make a change C to the master plan M iff the marginal expected-utility of C is positive, i.e., iff **expected-utility**($M\Delta C$) > **expected-utility**(M).

This is my proposal for a theory of rational decision making. I propose this principle as a replacement for classical decision theory and as a replacement for "simple" plan-based decision theory. It captures the basic insight that rational agents should guide their activities by considering the probabilities and utilities of the results of their actions, and it accommodates the observation that actions must often be selected as parts of plans, and the observation that optimality makes no sense for plans outside of toy examples. A decision-theoretic planner should be an evolutionary planner, not an optimizing planner. The principle of locally global planning tells us how an evolutionary planner should work. This proposal provides the starting point for building a decision-theoretic planner, because it tells us what we should want our implementations to accomplish. It involves a fundamental change of perspective from prior approaches to decision-theoretic planning, because decision-theoretic planning becomes a non-terminating process without a precise target rather than a terminating search for an optimal solution.

15. Incremental Decision-Theoretic Planning

This final section will make some first, tentative, steps towards implementation. At this point, they are highly speculative. The principle of locally global planning forms the basis for an "evolutionary" theory of rational decision making. It is a very general principle. An agent that directs its activities in accordance with this principle will implement the algorithm scheme diagrammed in figure 5. However, a complete theory of rational decision making must include more details. In particular, it must include an account of how candidate changes *C* are selected for consideration. This is an essential part of any theory of how rational agents should go about deciding what actions to perform. The selection of candidate changes is a complex matter and all I can do here is sketch a preliminary account, but it is important to at least do that in order to make it credible that locally global planning can provide the basis for rational decision making.

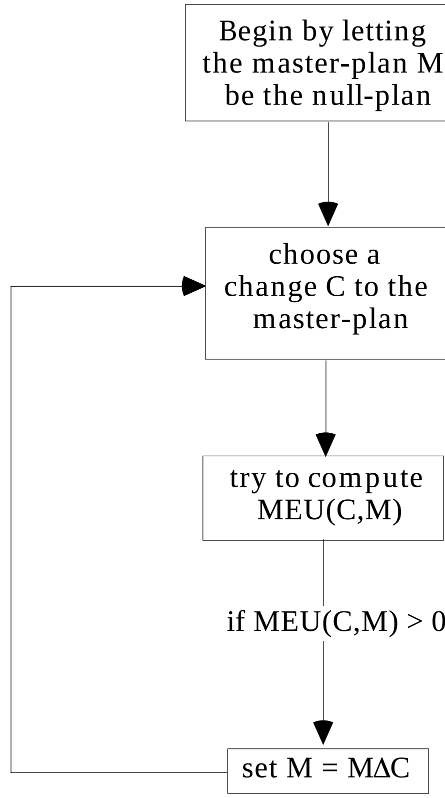

Figure 5. An algorithm schema for locally global planning

The logically simplest algorithm having the form of figure 5 would select potential changes randomly, evaluate them by computing their marginal expected-utilities, and on that basis decide whether to make them. This would be a version of the British Museum algorithm. But like any use of the British Museum algorithm, it is computationally infeasible. There are infinitely many possible changes. A feasible algorithm must select potential changes in some more intelligent fashion, concentrating on changes that have some reasonable chance of being desirable changes. Algorithms that direct the search for possible changes to the master plan are *planning algorithms*. Planning algorithms have been a traditional topic of research in AI, although most AI work on planning has considered only deterministic planning, where it is assumed to be known with certainty whether an action can be performed and what the results of performing it will be. Algorithms of interest in the present context will be algorithms for decision-theoretic planning. The details of constructing such an algorithm are going to be complex, and there is no hope of working them out completely in the space of this paper. However, some general remarks may be in order.

Let it be acknowledged immediately that there may be more than one good way of expanding the algorithm schema of figure 5. In deterministic planning, AI planning theory has produced a number of quite different planning algorithms, and the same thing may be true in decision-theoretic planning. The planning context and the cognitive resource bounds of the agent may determine which algorithms are best, with the result that there may be no single algorithm that is best for all agents and all times. So the remarks I will make here should be viewed as suggestions for one way in which decision-theoretic planning can work. This may not be the only way.

If we reflect upon human planning, it seems to have the general structure of what is called "refinement planning". In refinement planning, the planning agent begins by constructing a crude local plan, then searches for "flaws" in the plan, and refines the plan to remove the flaws. The flaws can be either internal to the plan or external to the plan. The external flaws derive from relations of "destructive interference" between the plan and the agent's master plan. Normally, refinement planning just refines the plan being constructed. In locally global planning, refinement planning must be generalized, allowing both the local plan and the master plan to be refined in light of the discovery of destructive interference between the two plans. This can produce multi-plan changes to the master plan. It is crucial to this general approach that the construction of multi-plan changes is driven by the construction of individual local plans. The multi-plan changes are not selected at random — they are changes involving newly constructed local plans and sets of previously adopted plans found to interfere with or be interfered with by the new local plan.

My proposal for how the master plan can be constructed and incrementally improved will be based on the account of refinement planning that I will sketch below. But before we do that, it is best to step back and ask why we should expect this general approach to planning to lead to the incremental improvement of the master plan. This can be justified if we make four defeasible assumptions. I will refer to these as the *pivotal planning assumptions*:

Assumption 1: The process of constructing "crude local plans" produces plans that will normally have positive expected values.

Assumption 2: Ordinarily, the expected value of the result of merging two plans will be the sum of the expected values of the two plans.

Assumption 3: Computationally feasible reasoning procedures will reveal those cases in which the second assumption fails.

Assumption 4: There will be "repair techniques" that can often be used to modify either the local plans or the master plan in such a way as to remove the destructive interference leading to the failure of the second assumption without having to replan from scratch.

The justification of these assumptions will unfold below. Given the pivotal planning assumptions, the planning agent can begin the construction of the master plan by constructing a single local plan having a positive expected value, and take that to be the master plan. Then the agent can systematically construct further local plans with positive expected values, and on the basis of the second assumption it can be assumed defeasibly that each time one of them is merged with the existing master plan, the result will be a master plan with a higher expected value. On the basis of the third assumption, rational investigation will enable the agent to discover those cases in which the defeasible assumptions fail. This amounts to discovering destructive interference. The fourth assumption tells us that it will often be possible to refine the local plan and / or the master plan so as to avoid the destructive interference, thus leading to a modification of the original plans which, when merged, produce a master plan with a higher expected value than the original master plan. By proceeding in this way, a rational agent can systematically evolve progressively better master plans.

Now let us turn to the justification of the pivotal planning assumptions. This will simultaneously generate a better understanding of the way in which refinement planning might work in order to make the four assumptions true.

15.1 Goal-Directed Planning

The evolution of the master plan begins with the construction of what I have called “crude local plans”. The first pivotal planning assumption is that rational cognition can produce plans that have some reasonable chance of having positive expected-utilities and hence of constituting useful additions to the master plan. We have little chance of finding such plans by constructing plans randomly and then assessing their expected-utilities. One way to achieve more intelligent plan construction is goal-directed planning. In goal-directed planning, the agent adopts goals, and then searches for ways of achieving them by using goal-regression (means-end reasoning). In goal-regression, we observe that a goal could probably be achieved in a certain way if a certain condition were satisfied. If the condition is not currently satisfied, we adopt the condition as a new subgoal. In this way, we work backwards from goals to subgoals until we arrive at subgoals that are already satisfied. The details of goal-regression planning are complex, but see my (1998) for a general theory of deterministic goal-regression planning.

Goal-regression is not the only game in town. See Weld (1999) for a discussion of some alternative approaches to goal-directed planning in AI. I am skeptical, however, of being able to apply the current “high-performance” planning algorithms in domains of real-world complexity. This is because of the algorithms’ stringent knowledge requirements — they make essential use of the closed world assumption, and that is totally unreasonable outside of toy problems.

Goals are simply valuable states of affairs — states of affairs that, were they achieved, would add to the overall value of the world (from the agent's point of view). Any valuable state of affairs can be chosen as a goal. In the cognitive architecture of a rational agent, that has the effect of initiating the search for plans for achieving the goal. The agent will probably not be in a position to construct plans for achieving most valuable states of affairs, so although they may technically be regarded as goals, they will have no effect on the agent's reasoning. They will simply be recorded as desirable, and then left alone unless the agent later encounters a way of achieving them.

The significance of goal-directed planning is that it produces local plans for the achievement of valuable goals. Assuming that the planning process will monitor the cumulative execution costs of the plan steps and not produce plans with high execution costs for lower-valued goals, goal-directed planning will produce plans that it is defeasibly reasonable to expect to have positive expected-utilities. Then in accordance with the second pivotal planning assumption, it is defeasibly reasonable for the agent to expect to be able to simply add the new plans to the master plan and thereby improve the master plan.

Much work in AI planning theory has focused on goal-directed planning. However, this is not adequate as a general account of planning. This is most easily seen by thinking more carefully about goals. As I have described them, goals are "goals of attainment". That is, they are valuable states of affairs that can, in principle be brought about, thereby attaining the goal. Rational agents operating in an environment of real world complexity may continually acquire new goals in response to the acquisition of new knowledge about their environment. If I learn that my favorite author has written a new book, I may acquire the goal of reading it. If I learn that a tiger is stalking me, I may acquire the goal of escaping from the tiger. If I learn that my fuel level is running low, I may acquire the goal of getting more fuel. These are goals that I cannot have until I acquire the requisite knowledge.

It is important to realize that these new goals are not simply instances of more general goals of attainment that I had all along. For example, my desire not to run out of fuel when I realize that I am in danger of doing so derives from a general desire not to run out of fuel. But that general desire cannot be represented as a goal of attainment. In particular, it is not the goal of *never running out of fuel*. If that were my goal, then if I ever did run out of fuel that goal would become henceforth unachievable regardless of anything I might do in the future, and so there would be no reason in the future for trying to avoid running out of fuel. My general desire is more like a *disposition* to form goals. As such, it is not itself the target of planning.[5]

[5] My general desire might be represented as the desire to run out of fuel as infrequently as possible, but that is not a goal of attainment — there is no point at which I can be said to have achieved that goal.

In different circumstances, I may be presented with different opportunities or exposed to different dangers. An opportunity is an opportunity to achieve an outcome having positive utility. This is accomplished by adopting the achievement of that outcome as a goal and then planning accordingly. Similarly, a danger is a danger that some outcome of negative utility will occur. That should inspire the agent to adopt the goal of preventing that outcome, and planning accordingly.

Planning for such newly acquired goals is not different in kind from planning for fixed goals, so this creates no problem for a theory of goal-directed planning. What is harder to handle, however, is the observation that we often plan ahead for "possible dangers" or "possible opportunities", without actually knowing that they will occur. For example, while trekking through tiger country, I may note that I might encounter a tiger, observe that if I do I will acquire the goal of not being eaten by him, and with that in mind I will plan to carry a gun. The important point here is that I am planning for a goal that I do not yet have, and I may begin the execution of the plan before acquiring the goal (that is, I will set out on my trek with a gun).

These are examples of what, outside of AI, is usually called "contingency planning". We plan for something that *might* happen, without knowing whether it actually will happen. If it does happen, then there will be a goal of the sort goal-directed planning normally aims at achieving. What makes it reasonable to begin planning before we actually acquire the information that generates the goals — just on the promise that we might acquire the information — is that such contingency plans can have positive expected-utilities even before we acquire the information. So in decision-theoretic planning, such contingency planning is desirable, but the problem is to fit it into the framework of our planning theory. The goals are hypothetical, but the plans are real in the sense that we may not only adopt them but also start executing them before acquiring the goals at which they aim. Boutelier *et al* (1999) make a similar observation, taking it to be a criticism of goal-directed planning.

The best way to understand contingency planning is to think more generally about the function that goals play in planning. The rational agent is trying to find plans with positive expected-utilities. Goals are of only instrumental value in this pursuit. On pain of computational infeasibility, we cannot use the British Museum algorithm and randomly survey plans until we find plans with high expected-utilities. We must direct our planning efforts in some more intelligent fashion, and goals provide a mechanism for doing that. Goals are states of high value, so there is a defeasible presumption that a plan that achieves a goal will have positive expected-utility. Hence goal-directed planning is a mechanism for finding plans with positive expected-values. But what these examples show is that goal-directed planning by itself will not always suffice for finding such plans. It is too restrictive to require that planning must always be initiated by actual goals. It must be possible to plan for hypothetical goals as well. Hypothetical goals will become goals *if* something is the case. Escaping from the tiger will become a goal if I learn there is a tiger, and acquiring more fuel will become a goal if I learn that I am running low on fuel. To initiate planning for these hypothetical goals, it should suffice to know that the antecedent is sufficiently probable to make a plan for achieving the goal have a positive expected-utility. Of course, the latter is not something we can be certain about until we actually have the plan, so the *initiation* of planning cannot require knowing that. What we can do is take the value of the goal in the context of the planning to be the value the goal would have were the antecedent true, discounted by the probability the antecedent will be true. Then we can engage in planning "as if" the goal were a real goal, and evaluate the plan in terms of the attenuated value of the goal. If the resulting plan has a sufficiently high marginal expected-utility, we can add it to the master plan and begin its execution before the goal becomes real (e.g., we can carry a gun or top off the fuel tank).

The preceding remarks constitute only the briefest sketch of some aspects of goal-directed planning, but hopefully they will point the way to a general theory of goal-directed planning that can be incorporated into a more mature theory of locally global planning. These remarks are intended as a sketchy defense of the first pivotal planning assumption.

15.2 Presumptively Additive Expected-Utilities

The second pivotal planning assumption was that, ordinarily, the expected-utility of the result of merging two plans will be the sum of the expected-utilities of the two plans. It follows from this that we can assume defeasibly that the marginal expected-utility of a plan is equal to its expected-utility. In other words, when we add a new local plan to the master plan, the result is to increase the expected-utility of the master plan by an amount equal to the expected-utility of the local plan. It is this assumption that makes it feasible to try to improve the master plan incrementally, by constructing local plans and adding them to it.

Let M be the master plan, and P be a local plan we are considering merging with M to produce the merged plan $M+P$. Both M and $M+P$ will normally be nonlinear plans. The expected-utility of a nonlinear plan was defined to be the minimum of the expected-utilities of its linearizations. A linearization L of $M+P$ will result from inserting the steps of a linearization P^* of P into a linearization M^* of M. The steps of P^* will not be dependent on any of the steps of M^*, so the expected-utility of L will be the expected-utility of M^* plus the expected-utility of P^* *in the context of L*. By the latter, I mean that the probabilities and utilities employed in computing the expected-utility of P^* in that context will all be conditional on the previous steps of L having been attempted.

If it is defeasibly reasonable to expect the utilities and probabilities of elements of P^* to remain unchanged when they are made conditional on the previous steps of *L* having been attempted, it will follow that it is defeasibly reasonable to expect that **expected-utility**(*L*) = **expected-utility**(*M**) + **expected-utility**(*P**), and hence **expected-utility**(*M*+*P*) = **expected-utility**(*M*) + **expected-utility**(*P*). This divides into two separate expectations — that the utilities remain unchanged, and that the probabilities remain unchanged. I have argued elsewhere that both of these expectations are defeasibly reasonable. I will briefly sketch the arguments for these claims.

Direct inference is the kind of inference involved in deriving single-case probabilities from general probabilities. E.g., it governs the kind of inference involved in inferring the probability that it will rain today from general probabilities of rain in different meteorological conditions. I proposed a general theory of direct inference in my (1990). In my (2002) I showed that the general principles underlying direct inference imply a defeasible presumption of statistical irrelevance for single-case probabilities:

(IR) For any P,Q,R, it is defeasibly reasonable to expect that $\mathrm{PROB}(P / Q\&R) = \mathrm{PROB}(P / Q)$.

This immediately implies an analogous principle of irrelevance for causal probability:

(CIR) For any P,Q,R, it is defeasibly reasonable to expect that $\mathrm{C\text{-}PROB}_A(P / Q\&R) = \mathrm{C\text{-}PROB}_A(P / Q)$.

This is exactly the principle we need to justify the defeasible assumption that the probabilities relevant to the computation of the expected-utility of a plan do not change when the plan is merged with the master plan. Of course, the identity in (CIR) will often fail in concrete cases, but the point of (CIR) is that it is reasonable to expect it to hold unless we have a definite reason for thinking otherwise.

Turning to utilities, there can be no logical guarantee that conditional utilities do not vary with context. However, if this happened too often, we would be unable to compute utility-measures for complex combinations of parameters on the basis of the utility-measures of the individual parameters or small combinations of them, and that in turn would make decision-theoretic reasoning impossible. There must be at least a defeasible presumption that for any state of affairs *P* and circumstances *C*, $\mathbf{U}(P) = \mathbf{U}(P / C)$. One way to defend this is to think of the utility $\mathbf{U}(P)$ associated with a state of affairs *P* as the value caused by the agent's being in those circumstances. On this conception of value, being in circumstances *P* literally causes the associated quantity of value to exist. This conception of value is discussed in much more detail in my (2001). It has the consequence that general principles for reasoning about values can be derived from analogous principles for reasoning about causation. I proposed the following as a general principle regarding causes:

Causal Irrelevance

If the agent has no reason to think otherwise, it is defeasibly reasonable to think that P does not cause Q.

In other words, we expect disparate events to be causally independent unless we have some concrete reason for thinking otherwise. Given the causal conception of value, this has the consequence that it is defeasibly reasonable to expect states of affairs to be value-neutral, i.e., to cause no change in the quantity of value produced by a situation. In other words, $\mathbf{U}(P/C) = \mathbf{U}(P)$. This is the assumption that we need regarding the utilities that play a role in computing **expected-utility**(M+P). These remarks are admittedly very sketchy, but for a more sustained defense of this reasoning, see my (2001).

The second pivotal planning assumption follows from these two principles for reasoning defeasibly about probabilities and utilities.

15.3 Finding and Repairing Decision-Theoretic Interference

The third pivotal planning assumption is that computationally feasible reasoning will reveal those cases in which the second assumption fails. The second assumption was that when two plans are merged into a single plan, the expected-utility of that composite plan will be the sum of the expected-utilities of the constituent plans. When this assumption holds, let us say that the plans exhibit *decision-theoretic independence*. *Decision-theoretic interference* is the failure of decision-theoretic independence. In adding local plans to the master plan, our defeasible assumption is one of decision-theoretic independence, so what is needed is tools for detecting decision-theoretic interference. Consider the ways in which decision-theoretic interference can arise.

The expected-utility of a plan is determined by various probabilities and utilities. The probabilities are (1) the probabilities that goals will be achieved, side effects will occur, or execution costs will be incurred if certain combinations of plan steps are attempted, and (2) the probabilities that the agent can attempt to perform a plan step if certain combinations of earlier plan steps are attempted. The utilities are the values of goals, side-effects, and execution costs conditional on certain combinations of plan steps having been attempted. Decision-theoretic interference arises from embedding the plan in a larger context (the master plan) in which some of these probabilities or utilities change. Without going into detail, it is clear that algorithms can be designed for searching for decision-theoretic interference by looking for constituents of the master plan that will change these probabilities and utilities. For example, if a plan relies upon the probability $\textsf{C-PROB}_A(goal\,/\,subgoal)$, we can search for “underminers” *P* such that $\textsf{C-PROB}_A(goal\,/\,subgoal\ \&\ P) \neq \textsf{C-PROB}_A(goal\,/\,subgoal)$ and such that the master plan contains a subplan that achieves *P* with some probability. This will be analogous to threat detection in deterministic goal-regression planning.

The fourth and final pivotal planning assumption is that we will often be able to make relatively small changes to local plans or the master plan to avoid any decision-theoretic interference that is detected. This is analogous to what occurs in classical refinement planning, and many of the same repair techniques will be applicable. The simplest will consist of adding ordering constraints to *M+P* to avoid interference. Another way of repairing the decision-theoretic interference is to add further steps that prevent the lowering of the probability. This is analogous to what is called *confrontation* in classical planning (Penberthy and Weld 1992). The upshot is that familiar ideas taken from classical planning will be applicable here as well. There may also be other ways of resolving decision-theoretic interference that do not correspond to techniques used in classical planning. That is a matter for further research. But this much is clear. The discovery of decision-theoretic interference need not cause us to reject our local plans altogether. We may instead be able to modify them in small ways to resolve the interference. Of course, this will not *always* be possible. Sometimes the interference will be irresolvable, and then we must reject one or more local plans and look for other ways of achieving some of our goals. However, none of this makes decision-theoretic planning intractable for realistic (resource bounded) rational agents.

16. Conclusions

We need a theory of decision-theoretic planning before we can implement. Without a theory, we don’t know what we want our implementation to do. The standard approach is what I have called “simple plan-based decision theory”, and is usually defended by waving casually in the direction of classical decision theory. However, simple plan-based decision theory has the potential to conflict with classical decision theory, which selects actions one at a time rather than in packages (plans). To defend decision-theoretic planning, we must first argue that classical decision theory is wrong. I began by raising three difficulties for classical decision theory. The first difficulty consisted of the set of familiar problems that give rise to causal decision theory, and I proposed a version of causal decision theory to handle these problems. The second difficulty was that the prescriptions of classical decision theory would only be reasonable if action omnipotence held. Action omnipotence fails in two ways. We cannot always perform actions when we try, and sometimes we cannot even try. This difficulty can be overcome by evaluating actions in terms of their expected-utilities rather than their expected-values. The third difficulty is that actions cannot be evaluated in isolation. We often have to choose plans rather than individual actions. This led to simple plan-based decision theory according to which it is rational to adopt a plan iff it is an optimal plan from a set of alternatives.

Simple plan-based decision theory fails, because outside of toy problems there is no way to define optimality in such a way that it is reasonable to expect there to be optimal plans. The problems for optimality stem from two sources. First, in domains of real-world complexity, reasoning is non-terminating, producing potentially infinitely many plans and making it impossible to know one has found an optimal plan even if optimal plans exist. Second, because plans can be extended indefinitely by merging them with other plans, there is no appropriate set of competitors to use in defining optimality.

A decision-theoretic planner should be an evolutionary planner rather than an optimizing planner. An evolutionary planner finds good plans, and replaces them by better plans as they are found. I have proposed the principle of *locally global planning* as the structure for an evolutionary planner. Locally global planning is based on two observations: (1) planning produces local plans; (2) local plans cannot be evaluated in isolation. They must be evaluated in the context of all the agent's other plans. This is accomplished by considering the contribution a local plan makes to the master plan. The result is that the objective of rational deliberation should be to find a good master plan and to be on the continual lookout for ways of improving the master plan. Planning becomes a non-terminating process rather than a terminating search for an optimal solution. The details remain to be worked out, but there are general logical reasons for expecting locally global planning to work, and for expecting it to be implementable using generalizations of classical planning techniques.

Appendix: Proofs

This appendix contains the proofs of theorems not proven in my (2002).

Theorem 6: $\mathbf{EV}(A \textit{ if } P) = \text{PROB}(P)\cdot\mathbf{EV}(A\,/\,P) + \text{PROB}(\sim P)\cdot\mathbf{EV}(\textit{nil}\,/\sim P)$.

Proof:

$\mathbf{EV}(A \textit{ if } P)$

$$= \sum_{O \in \mathcal{O}} \mathbf{U}(O) \cdot [\textsf{PROB}(P) \cdot \textsf{C-PROB}_{try\text{-}A}(O / A) + \textsf{PROB}(\sim P) \cdot \textsf{C-PROB}_{nil}(O / \sim P)]$$

$$= \textsf{PROB}(P) \cdot \sum_{O \in \mathcal{O}} \mathbf{U}(O) \cdot \textsf{C-PROB}_{try\text{-}A}(O / A) + \textsf{PROB}(\sim P) \cdot \sum_{O \in \mathcal{O}} \mathbf{U}(O) \cdot \textsf{C-PROB}_{nil}(O / \sim P)$$

$$= \textsf{PROB}(P) \cdot \mathbf{EV}(A / P) + \textsf{PROB}(\sim P) \cdot \mathbf{EV}(nil / \sim P). \ \blacksquare$$

We defined:

$$\mathbf{MEV}(A_i \textit{ if } C_i // A_1 \textit{ if } C_1, \ldots, A_{i-1} \textit{ if } C_{i-1}) = \mathbf{EV}(A_1 \textit{ if } C_1, \ldots, A_i \textit{ if } C_i) - \mathbf{EV}(A_1 \textit{ if } C_1, \ldots, A_{i-1} \textit{ if } C_{i-1}).$$

We have:

Theorem 11: $\mathbf{MEV}(A_1 \textit{ if } C_1, \ldots, A_k \textit{ if } C_k) = \sum_{1 \leq i \leq k} \mathbf{MEV}(A_i \textit{ if } C_i // A_1 \textit{ if } C_1, \ldots, A_{i-1} \textit{ if } C_{i-1})$.

Lemma 1: If for each $i \leq n$, P entails $\sim C_i$, then

$$\textsf{C-PROB}_{A_1 \textit{ if } C_1, \ldots, A_n \textit{ if } C_n}(O / P) = \textsf{PROB}(O / P).$$

Proof by induction.

$$\textsf{C-PROB}_{A_1 \textit{ if } C_1}(O / P) = \textsf{PROB}(C_1 / P) \cdot \textsf{C-PROB}_{A_1}(O / P \& C_1) + \textsf{PROB}(\sim C_1 / P) \cdot \textsf{PROB}(O / P \ \& \sim C_1).$$

As P entails $\sim C_1$, $\textsf{PROB}(C_1 / P) = 0$, $\textsf{PROB}(\sim C_1 / P) = 1$, and $\textsf{PROB}(O / P \& \sim C_1) = \textsf{PROB}(O / P)$, so $\textsf{C-PROB}_{A_1 \textit{ if } C_1}(O / P) = \textsf{PROB}(O / P)$.

Suppose the theorem holds for n. Then

$$\textsf{C-PROB}_{A_1 \textit{ if } C_1, \ldots, A_{n+1} \textit{ if } C_{n+1}}(O / P)$$

$$= \textsf{PROB}(C_1 / P) \cdot \sum_{B \in \mathcal{B}} \textsf{PROB}(B / P \& C_1) \cdot \textsf{C-PROB}_{A_2 \textit{ if } C_2, \ldots, A_{n+1} \textit{ if } C_{n+1}}(O / P \& A_1 \& B \& C_1)$$

$$+ \textsf{PROB}(\sim C_1 / P) \cdot \textsf{C-PROB}_{A_2 \textit{ if } C_2, \ldots, A_{n+1} \textit{ if } C_{n+1}} O / P \& \sim C_1)$$

$$= \text{C-PROB}_{A_2\ if\ C_2,...,A_{n+1}\ if\ C_{n+1}}O/P) = \text{PROB}(O/P).\ \blacksquare$$

Lemma 2: If for each $i \le n+1$, $\sim C_1$ entails $\sim C_i$, and $\sim C_i$ is consistent with P, and $A_1\&...\&A_n$ entails C_n, then:

$$\text{PROB}_{A_1\ if\ C_1,...,A_{n+1}\ if\ C_{n+1}}(O/P) - \text{PROB}_{A_1\ if\ C_1,...,A_n\ if\ C_n}(O/P)$$

$$= \text{PROB}(C_1/P)\cdot\text{PROB}(C_2/P\&A_1)\cdot...\cdot\text{PROB}(C_{n+1}/P\&A_1\&A_2\&...\&A_n)$$

$$\cdot[\text{PROB}_{A_1,...,A_{n+1}}(O/P) - \text{PROB}(O/P\&C_{n+1})].$$

Proof by induction:

$$\text{PROB}_{A_1\ if\ C_1}(O/P) - \text{PROB}(O/P)$$

$$= \text{PROB}(C_1/P)\cdot\text{PROB}_{A_1}(O/P) + \text{PROB}(\sim C_1/P)\cdot\text{PROB}(O/P\&\sim C_1) - \text{PROB}(O/P)$$

$$= \text{PROB}(C_1/P)\cdot[\text{PROB}_{A_1}(O/P) - \text{PROB}(O/P\&C_1)]$$

$$+ \text{PROB}(\sim C_1/P)\cdot[\text{PROB}(O/P\&\sim C_1) - \text{PROB}(O/P\&\sim C_1)]$$

$$= \text{PROB}(C_1/P)\cdot[\text{PROB}_{A_1}(O/P) - \text{PROB}(O/P\&C_1)].$$

Suppose the theorem holds for *n*. Then

$$\text{PROB}_{A_2\ if\ C_2,...,A_{n+1}\ if\ C_{n+1}}(O/P\&A_1) - \text{PROB}_{A_2\ if\ C_2,...,A_n\ if\ C_n}(O/P\&A_1)$$

$$= \text{PROB}(C_2/P\&A_1)\cdot\cdot...\cdot\text{PROB}(C_{n+1}/P\&A_1\&A_2\&A_2\&...\&A_{n+1})$$

$$\cdot[\text{PROB}_{A_2,...,A_{n+1}}(O/P\&A_1) - \text{PROB}(O/P\&A_1\&C_{n+1})].$$

So

$$\textbf{PROB}_{A_1 \text{ if } C_1,\ldots,A_{n+1} \text{ if } C_{n+1}}(O/P) - \textbf{PROB}_{A_1 \text{ if } C_1,\ldots,A_n \text{ if } C_n}(O/P)$$

$$= \textbf{PROB}(C_1/P)\cdot\textbf{PROB}_{A_2 \text{ if } C_2,\ldots,A_{n+1} \text{ if } C_{n+1}}(O/P\&A_1)$$

$$+ \textbf{PROB}(\sim C_1/P)\cdot\textbf{PROB}_{A_2 \text{ if } C_2,\ldots,A_{n+1} \text{ if } C_{n+1}}(O/P\&\sim C_1)$$

$$- \textbf{PROB}(C_1/P)\cdot\textbf{PROB}_{A_2 \text{ if } C_2,\ldots,A_n \text{ if } C_n}(O/P\&A_1)$$

$$- \textbf{PROB}(\sim C_1/P)\cdot\textbf{PROB}_{A_2 \text{ if } C_2,\ldots,A_n \text{ if } C_n}(O/P\&\sim C_1).$$

By lemma 1,

$$\textbf{PROB}(\sim C_1/P)\cdot\textbf{PROB}_{A_2 \text{ if } C_2,\ldots,A_{n+1} \text{ if } C_{n+1}}(O/P\&\sim C_1)$$

$$= \textbf{PROB}(\sim C_1/P)\cdot\textbf{PROB}_{A_2 \text{ if } C_2,\ldots,A_n \text{ if } C_n}(O/P\&\sim C_1),$$

so

$$\textbf{PROB}_{A_1 \text{ if } C_1,\ldots,A_{n+1} \text{ if } C_{n+1}}(O/P) - \textbf{PROB}_{A_1 \text{ if } C_1,\ldots,A_n \text{ if } C_n}(O/P)$$

$$= \textbf{PROB}(C_1/P)\cdot\textbf{PROB}_{A_2 \text{ if } C_2,\ldots,A_{n+1} \text{ if } C_{n+1}}(O/P\&A_1)$$

$$- \textbf{PROB}(C_1/P)\cdot\textbf{PROB}_{A_2 \text{ if } C_2,\ldots,A_n \text{ if } C_n}(O/P\&A_1)$$

$$= \textbf{PROB}(C_1/P)\cdot[\textbf{PROB}_{A_2 \text{ if } C_2,\ldots,A_{n+1} \text{ if } C_{n+1}}(O/P\&A_1) - \textbf{PROB}_{A_2 \text{ if } C_2,\ldots,A_n \text{ if } C_n}(O/P\&A_1)]$$

$$= \textbf{PROB}(C_1/P)\cdot\textbf{PROB}(C_2/P\&A_1)\cdot\ldots\cdot\textbf{PROB}(C_{n+1}/P\&A_1\&A_2\&\ldots\&A_n)$$

$$\cdot[\textbf{PROB}_{A_1,\ldots,A_{n+1}}(O/P) - \textbf{PROB}(O/P\&C_{n+1})]. \quad \blacksquare$$

Lemma 3: If the problem is classical and for each $i \leq n+1$, $\sim C_1$ entails $\sim C_i$, and $A_1\&\ldots\&A_n$ entails C_n, then:

$$\textbf{MEV}(A_{n+1} \text{ if } C_{n+1} // A_1 \text{ if } C_1,\ldots,A_n \text{ if } C_n)$$

$$= \textbf{PROB}(C_1)\cdot\textbf{PROB}(C_2 / C_1 \& A_1)\cdot\ldots\cdot\textbf{PROB}(C_{n+1} / A_1 \& A_2 \& \ldots \& A_n)$$

$$\cdot\left[\mathbf{EV}(A_1,\ldots,A_{n+1}) - \mathbf{EV}(A_1,\ldots,A_n / C_{n+1})\right].$$

Proof:

$$\mathbf{MEV}(A_{n+1} \textit{ if } C_{n+1} // A_1 \textit{ if } C_1,\ldots,A_n \textit{ if } C_n)$$

$$= \mathbf{EV}(A_1 \textit{ if } C_1,\ldots,A_{n+1} \textit{ if } C_{n+1}) - \mathbf{EV}(A_1 \textit{ if } C_1,\ldots,A_n \textit{ if } C_n)$$

$$= \sum_{O\in\mathcal{O}} \mathbf{U}(O)\cdot\left[\textbf{PROB}_{A_1 \textit{ if } C_1,\ldots,A_{n+1} \textit{ if } C_{n+1}}(O) - \textbf{PROB}_{A_1 \textit{ if } C_1,\ldots,A_n \textit{ if } C_n}(O)\right]$$

$$= \sum_{O\in\mathcal{O}} \mathbf{U}(O)\cdot\textbf{PROB}(C_1)\cdot\textbf{PROB}(C_2 / A_1)\cdot\ldots\cdot\textbf{PROB}(C_{n+1} / A_1 \& A_2 \& \ldots \& A_n)$$

$$\cdot\left[\textbf{PROB}_{A_1,\ldots,A_{n+1}}(O) - \textbf{PROB}(O / C_{n+1})\right]$$

$$= \textbf{PROB}(C_1)\cdot\textbf{PROB}(C_2 / C_1 \& A_1)\cdot\ldots\cdot\textbf{PROB}(C_{n+1} / A_1 \& A_2 \& \ldots \& A_n)$$

$$\cdot\left[\mathbf{EV}(A_1,\ldots,A_{n+1}) - \mathbf{EV}(A_1,\ldots,A_n / C_{n+1})\right].$$

Theorem 12: If the problem is classical:

$$\mathbf{MEV}(\textit{try-}A_{n+1} \textit{ if } (\textit{can-try-}A_{i+1} \,\&\, \textit{the agent has tried } A_1,\ldots,A_n)$$

$$// \textit{try-}A_1 \textit{ if can-try-}A_1\ldots,\textit{try-}A_n \textit{ if } (\textit{can-try-}A_i \,\&\, \textit{the agent has tried } A_1,\ldots,A_{n-1}))$$

$$= \textbf{PROB}(\textit{can-try-}A_1)\cdot\textbf{PROB}(\textit{can-try-}A_2 / \textit{try-}A_1)\cdot\ldots$$

$$\cdot\textbf{PROB}(\textit{can-try-}A_{n+1} / \textit{try-}A_1 \,\&\, \textit{try-}A_2 \,\&\ldots\&\, \textit{try-}A_n)$$

$$\cdot\left[\mathbf{EV}(try\text{-}A_1,\ldots,try\text{-}A_{n+1}) - \mathbf{EV}(try\text{-}A_1,\ldots,try\text{-}A_n / can\text{-}try\text{-}A_{n+1})\right].$$

Define:

$$\mathbf{expected\text{-}utility}(A_{n+1} / A_1,\ldots,A_n)$$

$$= \mathsf{PROB}(can\text{-}try\text{-}A_1)\cdot\mathsf{PROB}(can\text{-}try\text{-}A_2 / try\text{-}A_1)\cdot\ldots$$

$$\cdot\mathsf{PROB}(can\text{-}try\text{-}A_{n+1} / try\text{-}A_1 \,\&\, try\text{-}A_2 \,\&\ldots\&\, try\text{-}A_n)$$

$$\cdot\left[\mathbf{EV}(try\text{-}A_1,\ldots,try\text{-}A_{n+1}) - \mathbf{EV}(try\text{-}A_1,\ldots,try\text{-}A_n / can\text{-}try\text{-}A_{n+1})\right].$$

It follows that:

Theorem 13: If the problem is classical:

$$\mathbf{expected\text{-}utility}(\langle A_1,\ldots,A_n\rangle) = \sum_{1 \le i \le n} \mathbf{expected\text{-}utility}(A_i / A_1,\ldots,A_{i-1}).$$

References

Blythe, Jim, and Manuela Veloso

1997 "Analogical replay for efficient conditional planning", *AAAI9*.

Boutelier, Craig, Thomas Dean, and Steve Hanks

1999 "Decision theoretic planning: structural assumptions and computational leverage", *Journal of AI Research* **11**, 1-94.

Draper, Denise, Steve Hanks, and Daniel Weld

1994 "Probabilistic planning with information gathering and contingent execution", *Proceedings of AIPS94.*

Gibbard, Alan and William Harper

1978 "Counterfactuals and two kinds of expected value", in *Foundations and Applications of Decision Theory*, ed. C. A. Hooker, J. J. Leach and E. F. McClennen, Reidel, Dordrecht, 125-162.

Haddawy, Peter, and Steve Hanks

1990 “Issues in decision-theoretic planning: symbolic goals and numeric utilities”, *Proceedings of the DARPA Workshop on Innovative Approaches to Planning, Scheduling, and Control*, 48-58.

Jeffrey, Richard

1965 *The Logic of Decision*, McGraw-Hill, New York.

Kushmerick, N., Hanks, S., and Weld, D.

1995 “An algorithm for probabilistic planning”. *Artificial Intelligence* **76**, 239-286.

Lewis, David

1981 “Causal decision theory”, *Australasian Journal of Philosophy* **59**, 5-30.

Ngo, L., P. Haddawy, H. Nguyen

1998 “A modular structured approach to conditional decision-theoretic planning”, *AIPS’98*, 111-118.

Onder, Nilufer, and Martha Pollack

1997 “Contingency selection in plan generation”, *ECP97*.

1999 “Conditional, probabilistic planning: a unifying algorithm and effective search control mechanisms”, AAAI 99.

Onder, Nilufer, Martha Pollack, and John Horty

1998 “A unifying algorithm for conditional, probabilistic planning”, *AIPS1998 Workshop on Integrating Planning, Scheduling, and Execution in Dynamic and Uncertain Environments.*

Peot, M. A., and D. Smith

1992 “Conditional nonlinear planning”, *Proceedings of the First International Conference on Artificial Intelligence Planning Systems*, 189-197.

Penberthy, J. Scott, and Weld, Daniel

1992 “UCPOP: a sound, complete, partial order planner for ADL”. *Proceedings 3rd International Conference on Principles of Knowledge Representation and Reasoning*, 103-114.

Pollock, John

1990 *Nomic Probability and the Foundations of Induction*, Oxford University Press.

1992 “New foundations for practical reasoning”, *Minds and Machines*, **2**, 113-144.

1995 *Cognitive Carpentry*, MIT Press.

1998 “The logical foundations of goal-regression planning in autonomous agents”, *Artificial Intelligence* **106**, 267-335.

2001 “Evaluative cognition”, *Nous*, *Nous*, **35**, 325-364.

2001a “The uses of contingencies in decision-theoretic planning”, in preparation. Draft available at http://www.u.arizona.edu/~pollock.

2002 “Causal probability”, *Synthese*, *Synthese* **132** (2002), 143-185.

Pryor, Louise, and Gregg Collins

1996 “Planning for contingencies: a decision-based approach”, *Journal of Artificial Intelligence Research* **4**, 287-339.

Ramsey, Frank

1926 “Truth and probability”, in *The Foundations of Mathematics*, ed. R. B. Braithwaite. Paterson, NJ: Littlefield, Adams.

Russell, Stuart and Peter Norvig

1995 *Artificial Intelligence: a Modern Approach.* Prentice Hall: Upper Saddle River, N. J.

Savage, Leonard

1954 *The Foundations of Statistics*, Dover, New York.

Simon, Herbert

1955 “A behavioral model of rational choice”, *Qart. J. Economic*s, **59**, 99-118.

Skyrms, Brian

1980 *Causal Necessity*, Yale University Press, New Haven.

1982 “Causal decision theory”, *Journal of Philosophy* **79**, 695-711.

1984 *Pragmatics and Empiricism*, Yale University Press, New Haven.

Sobel, Howard

1978 *Probability, Chance, and Choice: A Theory of Rational Agency*, unpublished paper presented at a workshop on Pragmatism and Conditionals at the University of Western Ontario, May, 1978.

von Neumann, J., and Morgenstern, O.

1944 *Theory of Games and Economic Behavior.* New York: Wiley.

Warren, D.

1976 “Generating conditional plans and programs”, in *Proceedings of the Summer Conference and AI and Simulation of Behavior.*Weld, Daniel

Weld, Daniel

1999 “Recent advances in AI planning”, *AI Magazine* **20**, 55-68.

Williamson, Mike and Steve Hanks

1994 “Beyond symbolic goals: a step toward utility-directed planning”, *AAAI Spring Symposium on Decision-Theoretic Planning*, 272-278.