Logical Foundations for Decision-Theoretic Planning

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

Abstract

In decision-theoretic planning, plans are compared in terms of their expected-values. It has been supposed that this is justified by classical decision theory, but in fact this is incompatible with classical decision theory. Plan-based decision theory must be defended by showing that classical decision theory is wrong in ways that necessitate appealing to plans instead of actions in isolation. This paper proceeds by raising difficulties for classical decision theory and suggesting ways of meeting them which eventually lead to a kind of plan-based decision theory. The resulting plan-based decision theory differs in two way from more familiar versions proposed in the literature on decision theoretic planning. First, it evaluates plans in terms of their "expected utilities", which is different from the expected-value of executing the plan. Second, the objective is not to find optimal plans. It is argued that most plans cannot be compared by comparing their expected-utilities, and this makes optimality ill-defined. Instead, plans must be evaluated in terms of their contribution to the expected-utility of the agent's "master plan", which results from merging all of the agent's currently adopted plans into a single plan.

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 [1], [2], [3], [4], [5], [6], [7], [13]). However, the rush to implementation has proceeded without careful consideration of the theoretical foundations of such planning. At least two separate issues arise. First, how should the expected-value of a plan be defined? Second, how should plans be selected for adoption on the basis of their expected-values? The normal presumption in AI 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 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 classical decision theory entails the falsity of plan-based decision theory rather than supporting plan-based decision theory. And as I will show below (see also my [8] and [9]), 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 is that classical decision theory is subject to numerous theoretical difficulties, and the repair of these difficulties leads eventually to a properly formulated plan-based decision theory. I will begin by raising objections to the classical definition 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. Then 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. Optimality is not even well-defined outside of toy examples, so plan adoption must be based on somewhat different considerations.

2. The Expected-Utility of an Action

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 actions. Classical decision theory assumes 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 **PROB**(O/A) of each outcome conditional on the performance of each action. Finally, we assume a *utility-measure* 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:

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

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.

The simplest difficulty for classical decision theory is that it assumes that actions can be infallibly performed — *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 my goal is to win the Boston Marathon. I note that I could achieve this goal by running 100 mph for the duration of the race. No other action I might consider would have any better chance of winning the race, so in a context in which the only source of value derives from winning the race, this action is optimal and classical decision theory dictates that choosing to run 100 mph for the duration of the race is a rational decision. This, of course, would be a silly decision, because I cannot run 100 mph. Does classical decision theory really prescribe this? Yes, because it only looks at the probabilities that outcomes will result if actions are performed. I will consider three rather obvious rejoinders to this counterexample.

2.1 Possible Actions

The simplest response would be to protest that running 100 mph for the duration of the race is not a real alternative, because it is not something it is possible for me to do. So the suggestion is to restrict membership in A to actions that I can possibly perform. This does not solve the problem, however. Suppose I am a very good marathon runner. I *might* be able to run 12 mph for the duration of the race, although the probability of that is very small. I have never run that fast in a marathon before, but day-to-day fluctuations in one's physical state yield a probability of one in 10^{10} that I can do it. Thus this action cannot be excluded from A on the grounds that it is impossible for me to perform it. Hence classical decision theory prescribes choosing this action over watching the race on TV. But this would still be an unreasonable choice because it is incredibly unlikely that I will be able to do it.

2.2 High-Level Actions

Another way to object to this counterexample is to protest that *running 100 mph for the duration of the race* is either not an action or not the sort of action the theory is about. Consider the claim that it is not an action. This might be based upon the observation that to run 100 mph for the duration of the race I must perform a whole sequence of (simpler) actions over an extended period of time. However, this is true of virtually every action you can think of. Actions exhibit a continuous range of abstractness. Consider the actions *wiggle your finger, walk across the room, make a cup of coffee, vacation in Afghanistan, save the world.* Some of these, like *make a cup of coffee*, are typical of the kinds of actions that a theory of rational choice should concern. For example, classical decision theory is intended to capture reasoning like the following:

> Should I make a cup of coffee, or work in the garden? I would get more immediate gratification out of having a cup of coffee, but it would make me edgy later. If I work in the garden, I will sleep well tonight. So perhaps I should do the latter.

However, you can only make a cup of coffee by performing a whole sequence of simpler actions over an extended period of time.

There is an intuitive distinction between high-level actions and low-level actions, but the distinction is not a dichotomy. It is a continuum, and actions falling in the middle of the continuum are indisputably the sort of action a theory of rational choice should be about. So we cannot save classical decision theory by insisting that it only makes recommendations about low level actions.

2.3 Basic Actions

The problem we have noted for classical decision theory is that high-level actions can be difficult to perform, or even impossible to perform in the present circumstances, and that ought to be relevant to their decision-theoretic evaluation. It is tempting to try to avoid this difficulty by restricting the dictates of decision theory to low-level actions. Even if this were to work, it would not provide a fully adequate theory of rational choice, because the practical decisions we make generally involve choices between fairly high-level actions. But perhaps a theory of rational choice for high-level actions could be based upon a theory of rational choice for low-level actions, and classical decision theory might provide the latter.

Most actions are performed *by* performing other lower level actions. Basic actions are actions that can be performed without performing them by performing lower level actions. These are actions that are under direct control of the agent's

hardware, like *wiggle your finger*. I have heard it suggested that action omnipotence holds for basic actions, and so classical decision theory should be restricted to those. However, even basic actions are not infallibly performable. If your finger has been injected with a muscle paralyzer, you may not be able to wiggle your finger. This can be relevant to practical decisions. 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 make 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.

2.4 Trying

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. Rather than choosing between moving my left index finger and moving my right index finger, perhaps my choice should be viewed as one between trying to move my left index finger and trying to move my right index finger. That handles the preceding example nicely. Assuming that the probability is 1 that I can (under the present circumstances) move my left index finger if I try, then the expected-value of trying to do it is ten dollars, and the expected-value of trying to move my right index finger is one hundred dollars times the probability that I will be able to move it if I try. If that probability is greater than .1, then that is what I should do.

If action omnipotence holds for trying then we could base decision theory upon comparisons of the expected-values of tryings. Unfortunately, it is not true that we can always try. Suppose I show you a wooden block, then throw it in the 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 the previous 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. E.g., 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.

2.5 Expected-Utility

The dictates of classical decision theory would only be reasonable for actions that are infallibly performable, but there do not appear to be any. Rather than trying to repair classical decision theory by restricting its dictates to a special class of well-behaved actions, it seems we must reformulate the theory so that it applies to ordinary actions but takes account of the failure of action omnipotence. 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.

It is tempting to suppose that we should simply discount the expected-value of performing an action by the probability that we will be able to perform it if we try. That does not quite 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. That 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.

However, it is not adequate to assess actions just in terms of the expected-value of trying to perform them, because as we have seen, we may not be able to try to perform an action, and that is relevant to its assessment. 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. More generally, we may be uncertain whether we will be able to try to perform the action at the appropriate time. For example, consider a war game in which we plan to use an airplane to attack our opponent, but there is some possibility that our opponent will destroy the airplane before we can use it. 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

expected-value of trying to perform an action by the probability that, under the present circumstances, we can try to perform it. 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.

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 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. But 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 try to 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 try to perform the actions. The expected-value that will accrue specifically as a result of trying to perform an action *A* is the difference between the expected-value of trying to perform *A* and the 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. In this example it is this marginal expected-value that should be discounted by the probability of being able to try to perform the action.

In general, things are slightly more complex. We can have a case in which all of the value attached to *try-A* derives from its being the case that the agent *can-try-A*. Then there is no point to actually trying *A*. This is captured by considering the difference between the expected-value of trying to perform *A* and the expected-value of not trying to perform any of the alternative actions *given that the agent can try to perform A*.

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 given that one can try, discounted by the probability that we can try to do that:

$$\mathbf{EU}(A) = \mathbf{PROB}(\textit{can-try-A}) \cdot [\mathbf{EV}(\textit{try-A}) - \mathbf{EV}(\textit{nil}/\textit{can-try-A})].$$

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 these. My proposal is then 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 all of the above examples in a reasonable way.

2.6 Conditional Policies

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? In fact, it has an intuitive significance. I will now argue that 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*.

Decision theory has usually focused on choosing between alternative actions available to us *here and now*. A slight generalization of this problem is important in some contexts. We sometimes make *conditional decisions* about what to do if some condition *P* (predating the action) 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. *Doing A if P* is a *conditional policy*. The expected-value of the conditional policy *A if P* is just the expected-value of *A* discounted by the probability of *P* plus the expected-value of doing nothing discounted by the probability of ~*P*:

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

It can then be shown that $\mathbf{EU}(A) = \mathbf{MEV}(\textit{try-A if can-try-A})$ ($= \mathbf{EV}(\textit{try-A if can-try-A}) - \mathbf{EV}(\textit{nil})$). (The proof of this and other theorems can be found in my [11] or [12].) 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).

3. Plan-Based Decision Theory

Thus far I have raised one difficulty for classical decision theory and suggested a modification of the theory that resolves the difficulty. At this point I want to raise a different problem for classical decision theory. Classical decision theory has us focus on actions individually and choose them one at a time. However, choices can be more complicated than this. Sometimes we must choose actions in groups rather than individually. Suppose my objective is to transport

a ton of silver and a ton of gold from one location to another. I have a one-ton truck. I 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 I would risk damaging the truck springs. The actions I am considering 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-utility is that of transporting both on a single trip, even though that risks damaging the springs. However, if I have time to make two trips, that might be the better choice. That is, I 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 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.

What is important about the preceding example is that we cannot choose the group of actions by choosing the individual actions in it on the basis of their expected-utilities. The members of the group would not be chosen individually on their own strength. Rather, a pairwise comparison of actions would result in choosing the action of transporting both the gold and silver on a single trip, and that is the intuitively wrong choice. In this example, it is the group itself that is 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 what we call *plans*. The actions in a plan may be good actions to perform only because they are part of a good plan. How then do we choose plans? The obvious proposal is to construct a variant of classical decision theory by applying it to plans. The suggestion would be that we choose between competing plans in terms of their expected-utilities. This is the justification for decision-theoretic planning. Note that it does not justify decision-theoretic planning on the basis of classical decision theory. Rather, the justification turns upon a prior rejection of classical decision theory, and is proposed as a way of meeting theoretical difficulties that classical decision theory cannot handle.

4. The Expected-Utility of a Plan

If we are to evaluate plans in terms of their expected-utilities, we must first have a definition of expected-utility for plans. The considerations that led us to evaluate actions in terms of expected-utilities rather than expected-values apply equally to plans. We can define the expected-utility of a plan on analogy to the definition of expected-utility for actions:

EU(p) = **MEV**(trying to execute p if the agent can try to execute p).

But 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. For lack of space, I will focus exclusively on linear plans in this paper. 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 I endorse the following plan for making a cup of tea: (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.

Notice, however, that 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 "contingency nodes". However, not all plans will impose such conditions on the execution of their steps. If a 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 contingency nodes. 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 $A_1,...,A_{i-1}$, 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.

The most important consequence of this analysis of trying to execute a simple linear plan is that it enables us to compute the expected-utility of the plan in terms of the 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 marginal expected-value. Simple linear plans can also be viewed as conditional policies, although of a more complex sort. Let A_1 *if* $C_1,...,A_k$ *if* C_k be the *generalized conditional policy: do* A_1 *if* C_1, *then do* A_2 *if* C_2, *then* ... . The probability of an outcome on a generalized conditional policy can be expressed recursively as follows:

$$\mathbf{PROB}(O/(\sim C_1 \vee A_1)\ \&...\&\ (\sim C_k \vee A_k))$$
$$= \mathbf{PROB}(C_1)\cdot\mathbf{PROB}(O/A_1\&C_1\&(\sim C_2 \vee A_2)\ \&...\&\ (\sim C_k \vee A_k))$$
$$+\mathbf{PROB}(\sim C_1)\cdot\mathbf{PROB}(O/\sim C_1\&(\sim C_2 \vee A_2)\ \&...\&\ (\sim C_k \vee A_k)).$$

We can define conditional expected-values $\mathbf{EV}(\sigma/P)$ and *conditional marginal expected-values* $\mathbf{MEV}(\sigma/P)$ by conditionalizing all the probabilities on P. Then the following theorem can be proven:

$$\mathbf{MEV}(A_1\ \mathit{if}\ C_1,...,A_k\ \mathit{if}\ C_k)$$
$$= \sum_{1\leq i\leq k} \mathbf{MEV}(A_i\ \mathit{if}\ C_i/(\sim C_1 \vee A_1)\ \&...\&\ (\sim C_{i-1} \vee A_{i-1})).$$

For a simple linear plan, the corresponding policy is a sequence of steps of the form *try-*A_i *if* (*can-try-*A_i & *you have tried* $A_1,...,A_{i-1}$). The expected-utility of the plan is the marginal expected-value of this policy. If we define conditional expected-utilities as follows:

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

it can be shown that

$$\mathbf{MEV}(\mathit{try\text{-}}A_i\ \mathit{if}\ (\mathit{can\text{-}try\text{-}}A_i\ \&\ \mathit{the\ agent\ has\ tried}\ A_1,...,A_{i-1})$$
$$/(\sim\mathit{can\text{-}try\text{-}}A_1 \vee \mathit{try\text{-}}A_1)\ \&...\&\ (\sim(\mathit{can\text{-}try\text{-}}A_{i-1}\ \&\ \mathit{the\ agent\ has\ tried}\ A_1,...,A_{i-1}) \vee \mathit{try\text{-}}A_{i-1}))$$
$$= \mathbf{EU}(A_i/A_1\ \&...\&\ A_{i-1}).$$

Thus the marginal expected-value of a simple linear play is the sum of the expected-utilities of its steps each conditional on the previous steps having been attempted.

$$\mathbf{EU}(\langle A_1,...,A_n\rangle) = \sum_{1\leq i\leq n} \mathbf{EU}(A_i/A_1,...,A_{i-1}).$$

Simple linear plans are not really very interesting. To make these results useful they must be extended to more complex varieties of plans. This can be done in essentially the same way it was done for simple linear plans, but space precludes going into that here.

5. Choosing Between Plans

Plan-based decision theory proposes to choose plans in terms of their expected-utilities. The obvious suggestion is to apply classical decision theory directly to plans and say that a rationality dictates the choice of an optimal plan, i.e., one such that no competing plan has a higher expected-utility. 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*. However, it is clear that we often want to choose between plans that compete in much weaker ways. For example, if I am trying to decide what to cook for dinner, and I am considering roasting a chicken or barbecuing lamb chops, I *could* do both. The plans do not compete strongly. But nevertheless, they are in competition. I do not want to adopt both plans because I only want one dinner. One way to capture this is in terms 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-utility than one of the original plans. Cooking two dinners has a lower expected-utility than cooking just one because (1) the execution cost is essentially the sum of the execution costs of cooking either dinner alone, but (2) I can only eat one dinner, so the payoff from cooking two dinners is no greater than the payoff from cooking just one. It might be proposed, then, that two plans are competitors iff they compete weakly, and hence:

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

Now I want to propose a very fundamental problem for plan-based decision theory (see my [8]). Decision theory is formulated in terms of whether there exists (in logical space) a competing plan with a higher expected-utility. The problem is that there will almost always exist such a competing plan. To illustrate, suppose again that I am choosing between roasting chicken and barbecuing lamb chops for dinner. Suppose the former has the higher expected-utility. This implies that the plan of barbecuing lamb chops for dinner is not rationally adoptable, but it does not imply that the plan of roasting chicken for dinner is adoptable, because some other plan with a higher expected-utility 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 barbecuing lamb chops for dinner and then later going to a movie. This plan still competes with the plan of roasting chicken for dinner, but it has a higher expected-utility. Thus the plan of roasting chicken for dinner is not rationally adoptable. However, the competing plan is not rationally adoptable either, because it is trumped by the plan of roasting chicken for dinner and then later going to the same movie.

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 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 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. 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.

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 "maximal plans". These are plans prescribing what the agent should do for all the rest of its existence. Maximal plans cannot be extended by adding subplans for new goals. Because maximal plans include complete prescriptions for what to do for the rest of the agent's existence, any two maximal plans will make different prescriptions, and so will strongly compete. It seems initially quite plausible to suppose that maximal plans can be compared in terms of their expected-utilities, and that a maximal plan is rationally adoptable iff it is optimal, i.e., iff no other maximal plan has a higher expected-utility.

The problem with this way of salvaging plan-based decision theory is that in a domain of real-world complexity, it is computationally impossible for agents with realistic computational resource constraints to construct maximal plans. We cannot plan ahead now for all possible contingencies. All we can do is construct local plans for isolated goals, and as we have seen, they cannot be chosen on the basis of having higher expected-utilities than their competitors. Thus if plan-based decision theory is going to work, it must employ expected-utilities in some more sophisticated way.

6. Real Rationality and Ideal Rationality

Classical decision theory tries to do two things. First, it tries to characterize what is the "objectively best thing to do". Given the actual probabilities and utilities of outcomes, it attempts to pick out a unique action (or set of equally good actions) that is objectively best. These are the *optimal actions.* Second, it tells us that rationality requires an agent to choose what he rationally believes to be an optimal action out of all the alternatives available. The prescription to choose an optimal action can seem sensible as long as it is presupposed that the choice of which action to perform is made from a manageably small set of alternatives.

I have now argued that individual actions are not the most general objects of rational decision making. In general, we must choose plans directly and actions only indirectly because they are prescribed by rationally adopted plans. The suggestion was then to apply classical decision theory directly to plans:

> It is rational for an agent to decide to perform an action iff it is prescribed by a plan the agent rationally believes

to be optimal under the circumstances.

This would be reasonable if the agent could choose plans by surveying a small set of alternative plans and choosing the best, but he cannot. This is because plans are mathematical or logical constructions of unbounded complexity. However exactly competition between plans is to be defined, it seems pretty clear that in the real world there will be infinitely many competitors for any given plan. Choosing between plans would then involve evaluating and comparing infinitely many plans, which is impossible. Existing planners avoid this problem only by focusing on toy problems having bounded sets of actions, bounded sets of goals, and very limited complexity. Furthermore, as we have seen, in the real world, for every competitor there is likely to be another with a higher expected-utility, so optimality makes no sense. The upshot is that we cannot select optimal plans by surveying the set of all alternative plans. 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 choice in real agents operating in complex environments — must set different standards. I have argued that in a general setting there may be no optimal plans, but even if there are, rationality cannot require selecting optimal courses of action.

7. Locally Global Planning

Although optimality may not be achievable, there is no similar problem for understanding what it is for a course of action to be "good". 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 desirable 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-utility of the plan. Rationality must require agents to select "good" plans. Given more time to reason, good plans might be supplanted by better plans. However, reasoning is unending. There will be no point at which an agent has exhausted all possibilities in searching for plans. Despite this, the agent must take action. He cannot wait for the end of an unending search before deciding what to do, so his 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 (whatever that might mean).

The upshot is that planning in rational agents is an incremental process. The agent constructs and adopts plans, but the search never ends. New plans may conflict with previously adopted plans, leading either to the rejection of the new plans or the withdrawal of the previously adopted plans. The agent's plans "evolve" as time passes and more reasoning is performed. When the time comes to act, the agent directs his actions on the basis of the present set of adopted plans.

Let us look more closely at just how plans are adopted incrementally. 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? Plan-based classical decision theory would have us compare the new plan with all possible competitors, but that is impossible. 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-utility 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 them may affect the probabilities and the utilities employed in computing the expected-utility of the new plan. For example, if the new plan calls for the agent to perform a feat of strength, the probability of the agent being able to do that may be fairly high. But if other plans the agent has adopted will result in the agent being very tired at that time, then the probability of being able to perform the feat of strength 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 new plan has to be evaluated "in the context of the agent's other plans". It is not entirely clear what that means, however. 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. This is not something that is revealed by just computing the expected-utility of the new plan.

So how should the agent decide whether to adopt a new 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 simply the result of merging all of the agent's adopted plans into a single plan. The significance of master-plans is that, unlike ordinary plans, they can be compared in terms of their expected-utilities. The reason we could not compare arbitrary plans in terms of their expected-utilities was that different plans may be "intuitively incomparable". They may get their values by aiming at achieving different goals. Furthermore, a plan can aim at multiple goals, so it may have a higher expected-utility than a different plan just because it aims at more goals, not because it goes about achieving them in a better way (e.g., with lower execution costs or a higher probability of success). However, master-plans aim at all of the agent's goals simultaneously, so this is not a difficulty for comparing master-plans in terms of their expected-utilities. The expected-utility 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-utility. Equivalently, rationality dictates that if an agent is choosing between two master-plans, it should choose the one with the higher expected-utility.

If a master-plan has a positive expected-utility, that means it is better than doing nothing, so this is a reason for adopting it. However, it is only a defeasible reason. If a different master-plan with a higher expected-utility is found later, then it is a better master-plan and should be adopted in place of the first master-plan. The general theory of rational choice this produces is that an intention to perform an action is rationally acceptable iff it is part of a rationally adoptable master-plan. The dynamics of rational choice all concern the selection of master-plans. A master-plan is defeasibly adoptable iff it has a positive expected-utility, and it remains adoptable iff no master-plan with a higher expected-utility has been found. Rational deliberation is aimed at finding better and better master-plans. When it comes time to act, an agent's actions must be based on its current master-plan. This is despite the fact that unlimited further deliberation might always produce better and better master-plans. Real agents have finite cognitive powers, and must base their actions on the best master-plan they have been able to find to date. The ideal of basing actions on optimal master-plans is a logical fiction that is computationally impossible, and probably not even intelligible.

A master-plan is a *global plan* for achieving all of the agent's goals simultaneously (insofar as that is possible). If the only way we had of finding a master-plan with a higher expected-utility than our current one was to plan all over again from scratch and produce a new master-plan essentially unrelated to our 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 them incrementally. When trying to improve our master-plan, rather than throwing it out and starting over from scratch, what we must do is try to improve it piecemeal, leaving the bulk of it intact at any given time.

Normal planning processes produce *local plans*. These are small plans of limited scope, aiming at isolable goals. 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 we construct a new local plan, what we want to know is whether we can improve the master-plan by adding the local plan to it. Thus when we construct a new plan, 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-utility of the master-plan. Let us define the *marginal expected-utility* of the local plan P to be the difference its addition makes to the master-plan M:

$$\mathbf{MEU}(P,M) = \mathbf{EU}(M+P) - \mathbf{EU}(M).$$

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 adoptable. Thus local plans should be evaluated in terms of their marginal expected-utilities, not in terms of their expected-utilities simpliciter.

So, planning produces local plans, aimed at achieving local goals. But the local plans cannot be assessed in isolation. They must be assessed in terms of their contribution to the master-plan, which is the agent's current global plan. The master-plan is not a maximal plan, but it is the agent's current best approximation to a maximal plan. The theory of 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:

> 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 the new plan may only increase the expected-utility of the master-plan if we simultaneously delete conflicting plans. The second reason is that plans may have to be added in groups rather than individually. 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:

$$\mathbf{MEU}(C,M) = \mathbf{EU}(M\Delta C) - \mathbf{EU}(M).$$

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 $\mathbf{EU}(M\Delta C) > \mathbf{EU}(M)$.

This is my proposal for a theory of rational decision making. I propose this principle as a replacement for classical 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.

8. Conclusions

I have argued that classical decision theory must be rejected for two reasons. First, it flounders on a simplistic view of actions. 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 second difficulty is that in general actions cannot be chosen in isolation. They must be chosen as parts of plans.

We can compute expected-utilities for plans, but we cannot save classical decision theory by adopting a 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. The problems for such a plan-based decision theory are two-fold. First, for computational reasons we must confine our attention to local plans, but there is no apparent way to define "alternative" for local plans so that rational choice consists of choosing an optimal plan from a bounded list of alternatives. Second, even if there were optimal plans, real agents could not be required to find them, because that would require surveying and comparing infinitely many plans.

The upshot is that the only appropriate objects for evaluation in terms of expected-utilities are master-plans. However, the objective of rational decision making cannot be the construction of optimal master-plans. First, there is no reason to think that there will exist optimal master-plans. Second, even if there were optimal master-plans, the search for them would be computationally intractable. Real agents would not be able to find them. Instead, the objective of rational deliberation must be to find a good master-plan, and to be on the continual lookout for ways of making the master-plan better. Real agents will not be able to construct master-plans as the result of single planning exercises. The plans are too complex for that. The master-plan must instead be constructed incrementally, by engaging in local planning and then merging the local plans into the master-plan.

This is a theoretical account of rational decision making. It does not tell us how to build a decision-theoretic planner, but it does tell us what we should want our decision-theoretic planners to accomplish, and that is a necessary preliminary to building planners.

9. Acknowledgments

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

References

[1] Blythe, Jim, and Manuela Veloso, "Analogical replay for efficient conditional planning", *AAAI97*, 1997.

[2] Boutelier, Craig, Thomas Dean, and Steve Hanks, "Decision theoretic planning: structural assumptions and computational leverage", *Journal of AI Research* **11**, 1-94, 1999.

[3] Haddawy, Peter, and Steve Hanks, "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, 1990.

[4] Ngo, L., P. Haddawy, H. Nguyen, "A modular structured approach to conditional decision-theoretic planning", *AIPS'98*, 111-118, 1998.

[5] Onder, Niluger, and Martha Pollack, "Contingency selection in plan generation", *ECP97*, 1997.

[6] Onder, Niluger, and Martha Pollack, "Conditional, probabilistic planning: a unifying algorithm and effective search control mechanisms", AAAI 99, 1999.

[7] Onder, Niluger, Martha Pollack, and John Horty, "A unifying algorithm for conditional, probabilistic planning", *AIPS1998 Workshop on Integrating Planning, Scheduling, and Execution in Dynamic and Uncertain Environments*, 1998.

[8] Pollock, John, "New foundations for practical reasoning", *Minds and Machines*, **2**, 113-144, 1992.

[9] Pollock, John, *Cognitive Carpentry*, MIT Press, 1995.

[10] Pollock, John, "Locally global planning", *Proceedings of the Workshop on Decision-Theoretic Planning*, AIPS2000, 2000.

[11] Pollock, John, "Causal probability", in preparation, available at http://www.u.arizona. edu/~pollock.

[12] Pollock, John, "A theory of rational action", in preparation, available at http://www.u. arizona. \edu/ ~pollock.

[13] Williamson, Mike and Steve Hanks, "Beyond symbolic goals: a step toward utility-directed planning", *AAAI Spring Symposium on Decision-Theoretic Planning*, 272-278, 1994.