(version of 1/18/02)

# Rational Choice and Action Omnipotence

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

**Abstract**

Counterexamples are constructed for the theory of rational choice that results from a direct application of classical decision theory to ordinary actions. These counterexamples turn on the fact that an agent may be unable to perform an action, and may even be unable to try to perform an action. An alternative theory of rational choice is proposed that evaluates actions using a more complex measure, and then it is shown that this is equivalent to applying classical decision theory to "conditional policies" rather than ordinary actions.

## 1. Classical Decision Theory and the Optimality Prescription

A theory of rational choice is a theory of how an agent should, rationally, go about deciding what actions to perform at any given time. For example, I may want to decide whether to go to a movie this evening or stay home and read a book. The actions between which we want to choose are perfectly ordinary actions, and the presumption is that to make such a decision we should attend to the likely consequences of our decision. 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.

Philosophers often assume that the problem of rational choice is solved by classical decision theory. They take classical decision theory to prescribe choosing actions that maximize expected-values, and they tend to assume uncritically that this is the right prescription. Let us call this *the optimality prescription*. Ultimately, the defense of the optimality prescription must rest upon showing that it leads to the right choices in concrete examples. This paper presents intuitive counterexamples to this fundamental prescription, and goes on to argue that it should be replaced by a prescription that evaluates actions in terms of a more complex measure than expected-values. This turns out to be equivalent to applying classical decision theory to certain kinds of "conditional policies" rather than to ordinary actions.

I have found it difficult to frame this discussion in a way that does not elicit specious objections from decision theorists. The difficulty is that different decision theorists have different ideas about what classical decision theory says. 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*.

---

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

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 being true if the action is performed:

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

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. This is the formal representation of the optimality prescription.[1]

Different approaches to decision theory arrive at the optimality prescription in different ways. Savage lays down axioms governing rational choice, and derives the optimality prescription from the axioms. Von Neumann and Morgenstern lay down axioms governing preference between consequences, where the consequences include participation in lotteries. Jeffrey lays down axioms regarding preferences between states of affairs, and then identifies actions with those states of affairs one can make true. These differences are important, but what I want to focus on here is the optimality prescription that is common to all the different versions of classical decision theory. If classical decision theory is taken to provide a theory of rational choice, it does it by endorsing the optimality prescription.

Philosophers who are not decision theorists often assume uncritically that the "actions" the optimality prescription concerns are ordinary actions. In this they are certainly encouraged by the authors of the classical works. If we look at those works, there is no textual reason to doubt that these were the kinds of actions they concerned. Jeffrey uses examples of choosing whether to *buy a weekend admission ticket to the beach* or *pay admission daily*, whether to *arm with nuclear weapons* or *disarm*, and whether to *take a bottle of red wine to a dinner party* or *take a bottle of white wine*. Savage uses the example of making an omelet and deciding whether *to break a sixth egg into a bowl already containing five good eggs*, or *to break the sixth egg into a saucer and inspect it before adding it to the bowl of eggs*, or *to discard the sixth egg without inspection*. Savage *defines* an act to be a function attaching a consequence to each state of the world. This is objectionable in that it requires acts to have their consequences deterministically, but Anscombe and Aumann (1963) generalize this by taking the functions to assign probabilities to outcomes. Presumably, as philosophers, we do not want to literally identify acts with these functions but rather to say that the acts have such functions associated with them. But regardless, this imposes no restrictions on the kinds of acts being considered. Such functions can be associated with ordinary actions.

On reading the classical works, it certainly seems that the optimality prescription is to be applied to ordinary actions. But I have been surprised to discover (in conversation) that some contemporary decision theorists do not see things that way. They regard classical decision theory as an abstract mathematical theory, characterized by its axioms, and in principle immune from philosophical objections regarding its prescriptions. On this construal, to get any concrete prescriptions out of classical decision theory, we need a second theory that tells us how the first

---

[1] Note that this formulation of classical decision theory leaves open the type of probability employed. In particular, **PROB** can refer to some kind of "causal probability" of the sort discussed in causal decision theory. See Gibbard and Harper 1978; Sobel 1978; Skyrms 1980, 1982, 1984; Lewis 1981 for discussion of causal decision theory. See my (2002) for my own preferred version of causal decision theory. The problems discussed in this paper are orthogonal to those giving rise to causal decision theory.

theory is to be interpreted. In particular, we need a theory telling us what the "decision-theoretic actions" are. On this view, classical decision theory just tells us that *something* is to be evaluated in terms of its expected-value, and leaves open how that it to be used by a theory of rational choice. If we supplement classical decision theory with a theory about what decision-theoretic actions are, we get a theory of rational choice, but if philosophical objections are raised to the prescriptions of that theory of rational choice, they do not bear on decision theory — only on the theory of how it is to be interpreted.

I have no objection if this is the way one wants to understand classical decision theory. However, understood in this way, classical decision theory does not constitute a theory of rational choice. To get a theory of rational choice, we must conjoin classical decision theory with a second theory about what decision-theoretic actions are. This paper makes three points. The first is that we cannot solve the problem of rational choice by identifying decision-theoretic actions with ordinary actions. If we do, the optimality prescription makes intuitively unreasonable prescriptions. The second point is that we can avoid the intuitive counterexamples by replacing the optimality prescription with a prescription that evaluates ordinary actions in terms of a more complex measure than expected-value. The third point is that the resulting theory of rational choice is equivalent to applying the optimality prescription to something more complex than ordinary actions — what I will call "conditional policies". So if we adopt the abstract understanding of classical decision theory, we can regard the conclusion of this paper as proposing an alternative interpretation of classical decision theory wherein decision-theoretic actions are not actions at all, but rather conditional policies.

This paper will focus on one kind of difficulty for the optimality prescription (as applied to ordinary actions). The discussion will proceed within the framework of classical decision theory. This leaves some important questions unanswered. For example, where do the sets A of alternatives and O of outcomes come from, and what kind of probability is employed in the computation of expected-values? I have addressed these questions elsewhere (my 2001 and 2002), but in this paper I will try to remain as neutral as possible about their answers. The problem discussed here is orthogonal to these issues, arising for any version of the optimality prescription. So I will simply give the decision theorist his formulation of the decision problem. The solution I will propose for the present problem can then be applied to any version of decision theory and the optimality prescription.

## 2. Action Omnipotence

Our target is a theory of rational choice between ordinary actions like *going to a movie* or *staying home and reading a book*. If the optimality prescription is to provide a direct answer to the question of how to make such choices, we must take the decision-theoretic actions to which the optimality prescription applies to be ordinary actions. However, if we understand the optimality prescription in this way, as being concerned with choices between ordinary actions, it is subject to a simple difficulty. The problem is that the optimality prescription would only be reasonable for actions that can be performed infallibly. This is the assumption of *action omnipotence*. To see that this is indeed required to make the optimality prescription defensible, consider a simple counterexample based on the failure of the assumption. Suppose again that I am choosing between going to a movie or staying home and reading a book. I may decide that I would get more pleasure out of going to the movie, and so if that is the only source of value relevant to the decision, classical decision theory prescribes going to the movie. But now suppose my only way of going to the movie is to take a bus, and I know that there is talk of a bus strike. Suppose I believe that there is a 50% chance that there is a bus strike now going on, and so there is only a 50% chance that I will be able to go to the movie. This is surely relevant to my decision whether

to go to the movie, but it does not affect the expected-value of *going to the movie*. That expected-value is a weighted average of the values of the possible outcomes that will result if I actually do go, and takes no account of the possibility that I will be unable to go.

When faced with a problem like this, practical decision analysts simply reformulate the problem. The strategy is to take the uncertainty out of the action and move it into the consequences by taking the decision problem to be a choice between something like *trying to go to the movie* and *staying home and reading a book*. One might suppose that this should be regarded as a built-in constraint on classical decision theory — decision-theoretic actions must be infallibly performable. I have no objection if one wishes to stipulate this, although as remarked above it is hard to justify this on textual grounds. The examples used by the authors of the classical texts do not involve infallibly performable actions. They involve actions like *take a bottle of red wine to a dinner party*.

Suppose we do stipulate that the optimality prescription is only to be applied to infallibly performable actions. That avoids the counterexample, but remember that our topic is a theory of rational choice, and this concerns ordinary actions — not just infallibly performable actions. It takes little reflection to realize that no ordinary actions (of the sort Jeffrey and Savage used as examples) are infallibly performable. There are basically two ways a theory of rational choice might deal with the failure of action omnipotence. We can insulate the optimality prescription from the difficulties stemming from the failure of action omnipotence by restricting it to infallibly performable actions, but only at the cost of making it no longer provide a direct answer to questions of rational choice between ordinary actions. At the very least, it then needs to be supplemented with an account of how choices of ordinary actions are to be made by reference to choices between infallibly performable actions. Alternatively, we can look for a "decision-theory-like" theory of rational choice that applies directly to ordinary actions but modifies the way in which actions are evaluated. I will consider the former strategy first, but argue that it fails because there are no infallibly performable actions. I will take that to motivate the second strategy.

# 3. Restricting the Scope of the Optimality Prescription

The first suggestion is that although the optimality prescription makes incorrect prescriptions if we take it to apply to actions that are not infallibly performable, we can avoid this difficulty by restricting the range of actions to which the prescription is applied. I will argue that this simple strategy fails because there are no infallibly performable actions. I am unaware of any concrete proposals in the literature for what to take as the infallibly performable actions, so we are on our own in looking for candidates. I will consider four ways of trying to restrict the scope of the optimality prescription so as to avoid the problem of action omnipotence.

## 3.1 High-Level Actions

One way to reply to the counterexample given above is to protest that *going to the movie* 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 go to the movie I must perform a whole sequence of (simpler) actions over an extended period of time. Note, 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 Costa Rica, save the world.* At least 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, we want 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 among the actions a theory of rational choice should be about. So we cannot save classical decision theory as a theory of rational choice by insisting that it only makes recommendations about low-level actions.

## 3.2 Acts and Actions

To sort out the problems for applying classical decision theory to rational choice, we need a clearer understanding of actions. Let us begin with a type/token distinction. I will say that *acts* are individual spatio-temporally located performances (I mean to include mental acts here). I will take *actions* to be act-types. In rational decision making it is actions we are deliberating about. That is, we are deciding what type of act to perform.

The individuation of acts is philosophically problematic. If I type the letter "t" by moving my left index finger, are the acts of typing the letter "t" and moving my left index finger two acts or one? I don't think that this question has a predetermined answer. We need some legislation here. On the one hand, acts might be broadly individuated in terms of what physical or mental "movements" the agent makes. In this sense, I only performed one act.[2] However, some philosophers have insisted that two acts are performed in this case. One way to achieve this result is to take the type that is part of the specification of the act to be determined by the agent's intentions.[3] On this construal, if I type the letter "t" inadvertently when I move my finger, that is not an act that I performed. These are acts *narrowly individuated*. For present purposes, it is more convenient to take acts to be narrowly individuated, so that will be the convention adopted in this paper. If I want to talk about acts individuated broadly, I will say that explicitly. It is important to realize that this really is just a convention. If one takes broadly individuated acts to be basic, narrowly individuated acts can be identified with ordered pairs $\langle act, action \rangle$ where *act* is a broadly individuated act of type *action*.

As the above example illustrates, I often perform an act of one type *by* performing an act or sequence of acts of other types. Goldman (1970) called this "level generation". High level acts are performed by performing one or more lower level acts. E.g., I make a cup of coffee by performing the sequence of acts consisting of walking into the kitchen, putting water and coffee in the coffee pot, turning it on, waiting a while, and then pouring the brewed coffee into a cup. In turn, I put water in the coffee pot by picking it up, putting it under the tap, turning on the water, waiting until the water reaches the appropriate level in the pot, turning off the tap, and setting the coffee pot down on the counter. We can progress to lower and lower levels of acts in this way, but eventually we will reach acts like *grasp the handle of the coffee pot* or *raise my arm* that I can perform "directly" — without performing them by performing some simpler act.

As I am construing it, level generation is a relationship between acts, not actions. If one wants, one can define a corresponding relation between actions. You *can perform* an action *A by*

---

[2] This view is endorsed by Anscombe (1958), Davidson (1963), Schwayder (1965), and others.

[3] Danto (1963) and Goldman (1970) endorse the "two acts" theory, although their interest is in more than intentional acts, so they cannot include an intention in the act specification. My interest here is exclusively in intentional acts, because they alone are products of rational deliberation.

*performing* an action *B* just in case there can be an act of type *A* that is performed by performing an act of type *B*. However, I don't find this notion particularly useful.

The literature on action theory defines *basic acts* to be acts that are not performed by performing another act. For example, wiggling my finger will normally be a basic act. Note that the basic/non-basic distinction only makes sense for acts narrowly individuated — not for acts broadly individuated. If I type the letter "t" by moving my finger, there is only one broadly individuated act performed. Note also that the distinction between basic and nonbasic acts makes equally good sense when applied to artificial agents like robots. There are some acts they can perform directly — moving their effectors — and others they can only perform by performing some simpler acts.

A nonbasic act can be performed by performing another nonbasic act. E.g., you can turn on the light by throwing a switch. But you throw the switch by moving your finger, which is a basic act. So it follows that you also turn on the light by moving your finger (level-generation is transitive). On the assumption that there cannot be an infinite chain of level-generation, it follows that a nonbasic act is always performed by performing some basic act or sequence of basic acts.

We can think of actions (act types) as characterized by the range of sequences of basic acts by which acts of that type can be performed. This will generally be an open-ended range. For example, I can turn on the light by throwing a switch, but with sufficient ingenuity I can always think of new ways of turning on the light, e.g., Rube Goldberg devices, training my dog to rub against the switch, wiring in new power sources, etc.

It is important to keep in mind that the basic/non-basic distinction is a distinction between acts, not actions. We might try defining a *basic action* to be an action (an act type) for which every act of that type is a basic act. However, so-defined there are no physical basic actions. For example, *wiggling my finger* is normally a basic act. But if my finger is paralyzed, I might wiggle it with my other hand. In that case, I wiggle my finger by doing something else, so it is not a basic act. Similarly, if an undersea robot has a malfunctioning arm, it might move that arm with its other arm in order to use the grasper on the broken arm. A more useful notion is that of a *potentially basic action*, which is an act type that *can* have tokens that are basic acts. Given the narrow individuation of acts, most actions are not potentially basic actions. At least for humans, *fixing a cup of coffee* is not, but *wiggling my finger* is.

Most mental actions are not basic either. For example, multiplying 356 by 123 is something I do by performing a sequence of simpler multiplications. Perhaps I usually multiply 2 and 2 by performing a basic act, but I need not. I might perform it as sequential addition. So even that is only a potentially basic action. However, there are a few mental actions that plausibly cannot be performed by doing something else. For example, *recalling my mother's maiden name* might be a basic action. I can do things to cause me to recall my mother's maiden name, but I don't recall it *by* doing those things, at least not in the sense that recalling it is constituted by doing those other things. So there may be some basic mental actions.

## 3.3 Basic Actions and Action-Omnipotence

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. As remarked above, 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.

I have heard it suggested (not in print, but then, nothing has been suggested in print) that action omnipotence holds for basic actions, and so classical decision theory should be restricted to those. But as we have seen, there are no basic physical actions. And actions that are merely

potentially basic are not infallibly performable. If your finger has been injected with a muscle paralyzer, you may not be able to wiggle your finger. (You might be able to wiggle it by doing something else, e.g., moving it with your other hand, but if you are sufficiently constrained you may be unable to move it at all.) 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 restricting classical decision theory to potentially basic actions yields a theory that 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.

## 3.4 Deciding

Henry Kyburg recently suggested in conversation that the problem of action omnipotence may be handled by taking the actions to be evaluated decision-theoretically to be *decidings*. I suggested something similar in my (1995). It is unclear whether *deciding-to-A* is always infallibly performable. Terry Connolly recently raised the question (again in conversation) whether an addicted smoker can literally decide to stop smoking (as opposed to just giving lip service to the decision). A more decisive objection to this proposal is that *deciding-to-A* can have consequences that are not intuitively relevant to whether *A* is the action that ought rationally to be chosen. For instance, *my deciding to run for Congress* might make my wife angry and make me feel guilty, however my actually running for Congress might have neither of these consequences because my wife might get excited about it and I would see that it was clearly the morally right thing for me to do. Thus it seems inappropriate to evaluate the action of running for Congress in terms of the possible outcomes of deciding to run for Congress.

A different problem is that there is a slippage between *deciding-to-A* and *A-ing* that does not seem relevant to evaluating whether *A* is the action I ought to choose. I often decide to do something without ever getting around to doing it. For example, I may decide to go to the grocery store this afternoon but be continually distracted by other tasks. The afternoon may pass without my going, despite the fact that I never changed my mind. This will affect the expected-value of deciding to go, but does not seem relevant to whether I should go.

## 3.5 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. This is also suggested by a passing remark in Jeffrey (1985, pg. 83). 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 example of the paralyzed muscle 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. 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 choose to do.

At first glance, it seems that trying might be a basic action and infallibly performable. Then we could salvage the optimality prescription by saying that talk of choices between actions is really short for talk of choices between trying to perform actions. Or we could say that we are choosing between actions, but actions are to be evaluated in terms of the expected-values of trying to perform them. However, things are not so simple. Trying to perform an action is at least not usually a basic action. It is something one can do by doing something else — normally

something physical. For example, I may try to move a boulder by placing a pry bar under it and leaning on the pry bar.

If action omnipotence held for trying, it would make no difference whether trying is a basic action. In that case 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 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 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.[4] 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.

What exactly is it to try to perform an action? If *A* is an action that is not potentially basic, you try to perform *A* by trying to execute a plan for doing *A*. This plan may either be the result of explicit planning on your part, or the instantiation of a plan schema stored in memory (e.g., how to get home from the office), or the instantiation of "compiled in" procedural knowledge (e.g., how to ride a bicycle). One way in which you may be unable to try to do *A* is that, at the time you are supposed to act, you have no plan for doing *A*. (Of course, you need not have settled on a plan at the time you decide to do *A* — e.g., I may decide to fly to LA before deciding what flight to take.) Another way in which it can happen that you are unable to try is when your plan for *A*-ing requires certain resources, and when the time comes for executing it you do not have (or, perhaps, believe you do not have) the resources. E.g., I may not be able to try to paint the block because I do not have any paint. Finally, plans often contain epistemic contingencies. At the time you make the plan you may expect to have the requisite knowledge when you go to execute the plan, but if you do not then you may not even be able to try to *A*. For instance, you may be unable to try to paint the block because you do not know which one it is (in a pile of blocks), or you do not know where it is, or you do not know where the paint is.

For a potentially basic action *A*, you can either try to perform *A* by trying to instantiate a plan for doing *A*, or by trying to perform *A* in a "basic way". It is unclear to me whether you can always try to perform a potentially basic action in a "basic way". For example, if you know that your finger has been amputated, can you still try to wiggle it? This might be handled by saying that wiggling your finger is no longer a potentially basic action. But if that is right, rational choice seems to require you to have beliefs about whether your future actions are going to be potentially basic at the time you try to perform them. For instance, your finger may be intact now but you might have to worry about its being amputated before the time you plan to wiggle it.

It is important to realize that decisions to perform actions are always made in advance of the time the action is to be performed. It makes no literal sense to talk about deciding whether to do something *now*. If "now" literally means "at this very instant", then you are either already performing the action or refraining from performing it. It is too late to make a decision. In

---

[4] Note that trying is intensional. You might try to paint the block on the table red without realizing it is the block you thought destroyed, but then you have not tried to paint the block you thought destroyed.

rational choice, there must always be some time lag between the decision and the action. The time lag can be a matter of years, but even when it is a matter of seconds this has the consequence that we cannot be absolutely certain either that we will be able to perform the action or that we will be able to try to perform the action. At least often, your inability to perform an action may result from your epistemic state, but it is not your current epistemic state that determines whether you will be able to try — it is your epistemic state at the time of performing the action. So even if one were to suppose that we can be absolutely certain about our current epistemic state, it would not follow that we can be certain about whether we will be able to perform the action at some future time. A consequence of this is that *trying-to-A* is not infallibly performable, and so does not provide a candidate for the kinds of infallibly performable actions required to make the prescriptions of classical decision theory reasonable.

### 3.6 Infallibly Performable Actions

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. Let us say, somewhat more precisely than we did above, that an action *A* is infallibly performable iff **PROB**(*A* / *try-A*) = 1 and **PROB**(*can-try-A*) = 1.[5] The optimality prescription would only be reasonable for actions that are infallibly performable, but there do not appear to be any. It is worth noting that for a Bayesian like Skyrms (1980) who believes in strict coherence (the principle that only necessary truths have probability 1), these probabilities can never be 1, so for theoretical reasons such a theorist is committed to there being no infallibly performable actions. But even without such theoretical reasons, there do not appear to be any infallibly performable actions. If there are no infallibly performable actions, then restricting the optimality prescription to infallibly performable actions makes it vacuous, and hence unable to provide the basis for a theory of rational choice. It seems we must seek a different principle of rational choice that applies to ordinary actions but takes account of the failure of action omnipotence. Perhaps this can be done by changing the measure that is to be optimized in assessing actions. The next section explores this possibility.

## 4. Expected-Utility

How might we change the way in which we assess actions so that it accommodates the failure of action omnipotence? 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:

**EV**(*A*)·**PROB**(*A* / *try-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

---

[5] It may be wondered what kind of probability is being used here. I am trying to remain as neutral as possible on this issue. If one is a subjectivist, then the probability can be subjective probability. My own favored answer would be the "mixed physical/epistemic" probability of my (1990) or Pollock and Cruz (2000). Roughly, the mixed physical/epistemic probability of *P* is the objective probability of *P* conditional on all of the agent's epistemically justified beliefs. See the references for details.

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. These costs are incurred even if one tries to *A* but fails, so their contribution to the assessment of the action should not be discounted by the probability that the action will be performed successfully if it is attempted. To include all of these values and costs in our assessment of the action, we might look at the expected-value of trying to perform the action rather than the expected-value of actually performing it:

**EV**(*try-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.

However, more is relevant to the assessment of an action than the expected-value of trying 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. That might be handled by excluding it from the set A of alternatives. But 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 are considering using an airplane to attack our opponent, but there is some possibility that our opponent will destroy the airplane before we can use it. If the plane is destroyed, we will be unable to try to attack our opponent by using it. Clearly, 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:

**EV**(*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.

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. 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 1.

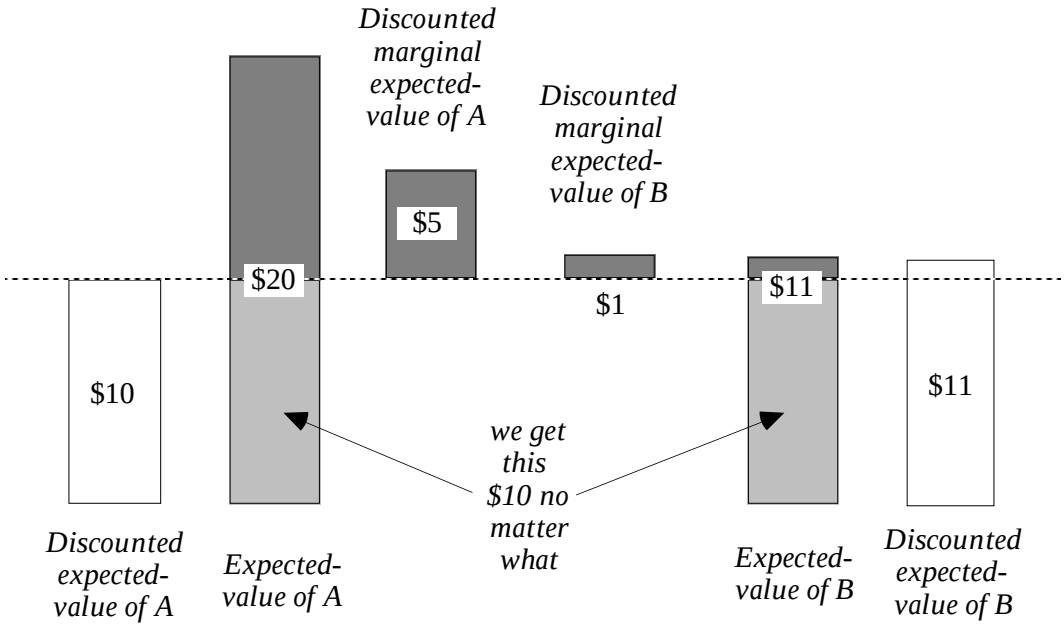


**Figure 1.** Discounted marginal expected-values.

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 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. It is this marginal expected-value that should be discounted by the probability of being able to try to perform the action.

A word about *nil*. I am trying to remain neutral about as many aspects of classical decision theory of possible. In particular, I do not want to make any unnecessary assumptions about where the set of alternative actions A comes from. Sometimes A will contain the null action as one of its members. Then *nil* is the action that is defined as not performing any of the other members of A. But I do not assume that *nil* will always be a member of A. That is not required for the way *nil* is used in computing marginal expected-values. There it is just used as a computational device — not as one of the alternatives in A. If *nil* is not included in A then it is defined as not performing any member of A.

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 to perform *A*. To illustrate, suppose you have a one-in-a-million chance of winning a lottery from which the winner receives one twenty dollar bill. Before the lottery, you have no money. Let *A* be the action of holding a twenty dollar bill in your hand and rotating it 360°. Let us suppose that this action has no cost, but it is also pointless — you get nothing out of doing it. Still, **EV**(*try-A*) is \$20, because in order to *try-A* you must have the twenty dollar bill. **EV**(*nil*) = $20 \cdot 10^{-6}$, which is effectively 0. Thus **EV**(*try-A*) – **EV**(*nil*) = \$20. But this does not mean that you should *try-A*. To avoid such spurious evaluations, we should instead consider 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*, i.e., **EV**(*try-A*) – **EV**(*nil* / *can-try-A*). By definition,

$$\mathbf{EV}(nil\,/\,can\text{-}try\text{-}A) = \sum_{O \in \mathrm{O}} \mathbf{U}(O)\cdot\mathbf{PROB}(O\,/\,nil\ \&\ can\text{-}try\text{-}A),$$

which in this case is \$20. Thus $\mathbf{EV}(try\text{-}A) - \mathbf{EV}(nil\,/\,can\text{-}try\text{-}A)$ is 0, reflecting the fact that there is no point in *trying-A*. In the general case, I will refer to the difference $\mathbf{EV}(try\text{-}A) - \mathbf{EV}(nil\,/\,can\text{-}try\text{-}A)$ as the *conditional marginal expected-value* of trying to perform the action.

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

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

If the probability that the agent can try to perform *A* is 0, the expected-value of trying to perform *A* is undefined, but in that case let us just stipulate that $\mathbf{expected\text{-}utility}(A) = 0$. 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 the optimality prescription by taking it to dictate choosing between alternative actions on the basis of their expected-utilities. With this change, the optimality prescription is able to handle all of the above examples in a reasonable way, without restricting it to a special class of particularly well-behaved actions.

However, this definition of the expected-utility of an action seems 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 on the basis of information available to us *here and now*. A slight 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*.[6] Conditional decisions are choices between conditional policies. We might regard (an extension of) classical 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:

$$\mathbf{PROB}_{A\ if\ P}(O) = \mathbf{PROB}(P)\cdot\mathbf{PROB}(O\,/\,P\&A) + \mathbf{PROB}(\sim P)\cdot\mathbf{PROB}(O/\sim P\&nil).$$[7]

---

[6] One can generalize conditional policies in various ways, e.g., looking at policies of the form "*Do A if P and do B if ~P*". Some of these generalizations are discussed in my (2002). However, we only need this simple form of conditional policy for the present discussion.

[7] Conditional policies are reminiscent of the "mixed acts" of Savage (1958), but Savage requires the probability of the outcome to be independent of *P* in a mixed act.

Then we can define straightforwardly:

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

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

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

This is the expected-value of the action given the assumption that $P$ is true. $\mathbf{EV}(A \textit{ if } P)$, 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 1:** $\mathbf{EV}(A \textit{ if } P) = \mathbf{PROB}(P) \cdot \mathbf{EV}(A / P) + \mathbf{PROB}(\sim P) \cdot \mathbf{EV}(\textit{nil} / \sim P)$.

Proof:

$$\begin{aligned}
\mathbf{EV}(A \textit{ if } P) &= \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot \mathbf{PROB}_{A \textit{ if } P}(O) \\
&= \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot [\mathbf{PROB}(P) \cdot \mathbf{PROB}(O / P \& A) + \mathbf{PROB}(\sim P) \cdot \mathbf{PROB}(O / \sim P \& \textit{nil})] \\
&= \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot \mathbf{PROB}(P) \cdot \mathbf{PROB}(O / P \& A) + \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot \mathbf{PROB}(\sim P) \cdot \mathbf{PROB}(O / \sim P \& \textit{nil}) \\
&= \mathbf{PROB}(P) \cdot \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot \mathbf{PROB}(O / P \& A) + \mathbf{PROB}(\sim P) \cdot \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot \mathbf{PROB}(O / \sim \& \textit{nil} P) \\
&= \mathbf{PROB}(P) \cdot \mathbf{EV}(A / P) + \mathbf{PROB}(\sim P) \cdot \mathbf{EV}(\textit{nil} / \sim P). \blacksquare
\end{aligned}$$

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.[8] 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:

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

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

---

[8] This is the "official" definition, e.g., above and in Savage (1953), Jeffrey (1965), etc. However, in classical decision theory values can be scaled linearly without affecting the comparison of expected-values, so in actual practice people generally employ marginal expected-values in place of expected-values. For comparing expected-utilities (defined as in this paper), such scaling can change the comparisons, so we must use marginal expected-values rather than expected-values.

**MEV**(*A* / *P*) = **EV**(*A* / *P*) – **EV**(*nil* / *P*).

**MEV**(*A if P*) = **EV**(*A if P*) – **EV**(*nil 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:

**MEV**(*A if P*) = **EV**(*A if P*) – **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 2: MEV**(*A if P*) > **MEV**(*B if Q*) iff **EV**(*A if P*) > **EV**(*B if Q*).

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

**Theorem 3: MEV**(*A if P*) = **PROB**(*P*)·**MEV**(*A* / *P*)

Proof:

**MEV**(*A if P*)

= **EV**(*A if P*) – **EV**(*nil if P*)

= **PROB**(*P*)·**EV**(*A* / *P*) + **PROB**(~*P*)·**EV**(*nil* / ~*P*)

– **PROB**(*P*)·**EV**(*nil* / *P*) – **PROB**(~*P*)·**EV**(*nil* / ~*P*)

= **PROB**(*P*)·**EV**(*A* / *P*) – **PROB**(*P*)·**EV**(*nil* / *P*)

= **PROB**(*P*)·[**EV**(*A* / *P*) – **EV**(*nil* / *P*)]

= **PROB**(*P*)·**MEV**(*A* / *P*). ■

We defined **expected-utility**(*A*) = **PROB**(*can-try-A*)·[**EV**(*try-A*) – **EV**(*nil* / *can-try-A*)]. Because *A* entails *can-try-A*, **PROB**(*O* / *A* & *can-try-A*) = **PROB**(*O* / *A*), and hence **EV**(*try-A* / *can-try-A*) = **EV**(*try-A*). It then follows from theorem 3 that:

**Theorem 4: expected-utility**(*A*) = **MEV**(*try-A if can-try-A*).

Hence by virtue of theorem 2:

**Theorem 5: expected-utility**(*A*) > **expected-utility**(*B*)
iff **EV**(*try-A if can-try-A*) > **EV**(*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.

Note that if we adopt the "abstract" reading of classical decision theory according to which it is left open what count as decision-theoretic actions, then one way of understanding the present proposal is that the optimality prescription yields the correct prescriptions if it is restricted to these conditional policies. On the other hand, most variants of classical decision theory derive the optimality prescription from general axioms about preferences. It would require a careful examination of those theories to see whether their axioms remain reasonable when decision-theoretic actions are identified with these conditional policies. That is a matter that I will not pursue here.

# 6. Conclusions

A theory of rational choice is a theory of how an agent should, rationally, go about deciding what actions to perform at any given time. This is a theory about choosing between ordinary actions. If we adopt the optimality prescription of classical decision theory as our theory of rational choice, taking the actions to which it applies to be ordinary actions, we find that it can yield intuitively incorrect decisions by ignoring the fact that it is sometimes difficult or impossible to perform an action and sometimes one cannot even to try to perform it. One way to attempt to repair the theory is to restrict it to a class of actions that can be performed infallibly, but there does not seem to be any appropriate class of actions having this property. The only other alternative appears to be to revise the definition of expected-value to take account of the failure of action omnipotence. That led us to the rather ad hoc looking definition of "expected-utility". However, that definition can be made intuitive by the observation that comparing actions in terms of their expected-utilities is equivalent to comparing them in terms of the expected-values of the associated conditional policies of the form *try-A if can-try-A*. In light of the failure of action omnipotence, choosing an action is tantamount to adopting such a conditional policy, and so the evaluation of the action should be the same as the evaluation of the conditional policy.

Let me close by remarking that I have suggested a way of repairing one difficulty for classical decision theory. I would stop far short of claiming that the repaired theory constitutes a satisfactory theory of rational choice. It is my conviction that a number of other serious problems remain. I intend to address them elsewhere.[9]

# References

Anscombe, G. E. M.
1958 *Intention*. Ithaca: Cornell University Press.
Anscombe, F. J. and R. J. Aumann
1963 "A definition of subjective probability", *Annals of Mathematical Statistics* **34**, 199-205.
Danto, Arthur
1963 "What can we do?", *The Journal of Philosophy*, LX, 435-445.
Davidson, Donald
1963 "Actions, reasons, and causes", *The Journal of Philosophy*, LX.
Gibbard, Alan and William Harper

[9] See particularly my (2001).

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.

Goldman, Alvin

1970 *A Theory of Human Action*, Princeton University Press.

Jeffrey, Richard

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

1983 *The Logic of Decision, 2nd edition*, Chicago University Press, Chicago.

Lewis, David

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

Pollock, John

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

1995 *Cognitive Carpentry*, MIT Press.

2001 "Plans and decisions", in preparation. Available at http://www.u.arizona.edu/~pollock.

2002 "Causal probability", *Synthese*, forthcoming. Available at http://www.u.arizona.edu/~pollock.

Pollock, John, and Joseph Cruz

2000 *Contemporary Theories of Knowledge*, Rowman and Littlefield.

Ramsey, Frank

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

Savage, Leonard

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

Schwayder, David

1965 *The Stratification of Behavior*, New York: Humanities Press.

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.