ABSTRACT. Counterexamples are constructed for classical decision theory, turning on the fact that actions must often be chosen in groups rather than individually, i.e., the objects of rational choice are plans. It is argued that there is no way to define optimality for plans that makes the finding of optimal plans the desideratum of rational decision-making. An alternative called “locally global planning” is proposed as a replacement for classical decision theory. Decision-making becomes a non-terminating process without a precise target rather than a terminating search for an optimal solution.

KEY WORDS. AI, alternatives, optimality, plans, rationality.

1. THE OPTIMALITY PRESCRIPTION

How should I go about making rational decisions? That is the fundamental question of the theory of rational choice. Human beings, and any real cognitive agents, are subject to cognitive resource constraints. They have limited reasoning power, in the form of limited computational capacity and limited computational speed. This makes it impossible, for example, for them to survey all of the logical consequences of their beliefs, or to compare infinitely many alternatives. This is a fundamental computational fact about real agents in the real world, and I would suppose that it could not have been otherwise. An account of how a real agent should make decisions must take account of these limitations.

Theories of rational decision-making are sometimes taken to be theories about how ideal agents, immune to such cognitive limitations, should make decisions (Cherniak, 1986; Lewis, 1981; Skyrms, 1980, 1984). One can, of course, choose to talk that way, but it is hard to see what that has to do with what we, as fallible human beings, should do. For instance, if a theory of
ideal agents says that they should attend to all of the logical consequences of their beliefs, but we as human beings cannot do that, then the recommendations applicable to ideal agents are simply not applicable to us. *We* should do something else. As I use the term “the theory of rational decision-making”, it is about what we, and other resource bounded cognitive agents, should do. I want to know how, given our cognitive limitations, we should decide what actions to perform.

In other words, I want a theory of *real rationality* as opposed to a theory of *ideal rationality*. This distinction is widely recognized, but it often seems to be supposed that as philosophers our interest should be in ideal rationality. The rationality a human can achieve is mere “bounded rationality”—a crude approximation to ideal rationality. But we come to the study of rational decision-making with an initial interest in how we, and agents like us, should make decisions. This is the notion of rationality that first interests us, and it is the target of this paper.

Decisions must often be made in the face of uncertainty regarding both the agent’s initial situation and the consequences of his actions. Most discussions of rational choice proceed against the background of classical decision theory, which is generally assumed uncritically. The basic ideas of classical decision theory can be stated simply. We assume that our task is to choose an action from a set $A$ of *alternative actions*. The actions are to be evaluated in terms of their outcomes. We assume that the *possible outcomes* of performing these actions are partitioned into a set $O$ of pairwise exclusive and jointly exhaustive outcomes. We further assume that we know the probability $\text{PROB}(O/A)$ 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-utility* 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. The crux of classical decision theory is that actions are to be compared in terms of their expected-utilities, and rationality dictates choosing an action that is *optimal*, i.e., such