Thinking about Acting: Logical Foundations for Rational Decision Making, by John L. Pollock. Oxford: Oxford University Press, 2006. Pp. 267. H/b £23.99.

This stimulating and challenging book describes ways in which limited agents should make decisions. As Pollock puts it, the question is: 'what should I do given that I am who I am, with all my actual cognitive limitations?' This under-describes the project. We have to say what kinds of limitations are relevant and what the 'should' amounts to. Pollock is concerned with limitations of memory and calculating power. And the normativity in question amounts to asking how agents subject to these limitations could be optimally designed. He is not concerned with issues of whether a good procedure for well-constructed limited agents might be impractical for actual human beings, or whether a description of a good way to think given one's limitations could be the basis for useful advice for human beings as they are. In this review I will ignore many of the AI aspects of the book, and thus many impressive analyses and persuasive suggestions, and concentrate on the question of decision-making procedures that it is in the interest of limited agents to adopt.

In early chapters Pollock presents a picture of how we actually make decisions. There are two main aspects. On the desire side of things people have 'state likings' and 'feature likings': they give favourable ranking to aspects of situation-types, which they try to bring about; and to particular qualities of their immediate experience, which in general they would like to get more of. To a first approximation, state liking is pleasure and feature liking is preference over coarse-grained possible worlds. The former of these should generate the second, and Pollock gives a plausible sketch of a kind of practical induction by which a person could figure out the causal connections between aspects of her environment in order to rank features in terms of their propensities to produce liked states. I am sure these routines will work better in some environments than in others, so an important question about Pollock's system is whether the actual normal environments of human beings are suitable to it.

On the belief side of things Pollock's approach is marked by distrust of subjective probability and the use of his own special objective probability, nomic probability, which amounts to the strength of the law-like connection between two attributes. Individuals have beliefs about nomic probabilities, so that a person can assign a number to the nomic probability that an attribute A is causally associated with an attribute B. Pollock would probably not like saying it this way, but this number is essentially the person's subjective probability that a randomly chosen A will be B. Pollock produces elegant variations on lottery considerations to argue that belief and subjective probability cannot be

combined, and resolves the tension by sticking with belief. So his agents have full beliefs, many of which are about causal properties of things, and use them in conjunction with causally oriented feature likings to choose actions. They do a lot of causal thinking.

Feature likings and beliefs combine to select actions. The primary objects of selection are not simple action-descriptions but plans, and they are chosen not in terms of their simple expected value but in terms of their marginal value in comparison with alternatives conditional on one's being able to carry them out. These two complications are responses to real and inevitable aspects of finite agents, and are discussed in chapters nine to twelve, the most interesting and radical parts of the book. Pollock discusses in illuminating detail problems arising from the fact that there is no point choosing an action which would have good results if performed, when you cannot perform it. In more interesting cases it is uncertain whether you can perform it. After a suggestive and imaginative discussion of the issue Pollock opts for evaluating acts in terms of their outcomes, conditional on the agent's being able to try to perform them. 'Try' is of course a very subtle and ambiguous word, but Pollock argues that some, at any rate, of the ambiguities are not going to affect the evaluation of the differences in value between actions, which is what ultimately drives choice.

Actions can be evaluated relative to other actions, but the results are often not very helpful since the value of doing A is usually very different from that of doing A & B. So what Bs are you to include with your A? Since the most stable objects of consequentialist evaluation are large (A & B & C & …), Pollock goes for structured plans as the things you should evaluate. In fact it seems that when deciding which paperback book to buy in an airport one should, on Pollock's account, first see each as part of a structured plan involving a number of acts linked by means–end connections, and then compare them together with some alternatives in the context of an even wider outline plan for a large chunk of one's life.

Notice what has happened. It is reminiscent of a theme of the philosophy of mathematics between roughly 1930 and 1960. When we are forced to do without the idealising assumption of infinite capacity we find that our procedures (whether they be mathematical proof or practical decision) become much more complicated. In a way, the finite is more complex than the infinite. So the rational agent who is building her limitations explicitly into her decision-procedures may face more rather than fewer complications of her thinking: enough to overwhelm her limited capacities. In Pollock's system the formal thinking that an agent has to do to accommodate the fact that she does not know what acts she can perform and does not know what plans are compatible with what other plans are more complicated than those required by the standard idealisations. It is hard enough being very finite; being knowingly finite seems harder yet.

There are several ways in which this is unfair to Pollock. Whenever he intro-

duces a new twist in the requirements on rational agents, he goes out of his way to argue that the computational burden is manageable. And after the long discussion of plans as objects of choice he acknowledges that long-term plans are diffuse things whose pros and cons are very hard to compare. So instead of requiring them to be subject to any formal value-giving procedure he presents a picture of a slowly evolving and very incomplete plan of life, in the context of which smaller plans are evaluated relative to the agent's capacity to carry them out. The overall plan is adopted on pragmatic grounds: one does not optimize with respect to it but sticks with it as long as it is serving its purpose and can be expected to produce good enough results. So wrapped around the rather involved optimization is something vaguely like satisficing.

This combination of aspects is not obviously wrong. The general strategy is to think as precisely as one can, taking into account as much as possible of the limitations of one's information and one's depth of thinking, directing the process in a general motivational direction that one does not subject to intense scrutiny. There is obviously a balance for any finite agent to strike between evaluating actions, given large-scale plans, and evaluating those plans. Pollock opts for putting a lot of juice into the detailed thinking, and rather less juice, or at any rate rather less meticulousness, into the overall direction. This may often be a good style. It is not an objection that it allows a finely-calculated voyage to what better navigation would have shown to be disaster. Any compromise between these unattainables will bring its own version of this. But it is not obvious that it is the best compromise. It is far from obvious that there is a best compromise. So the main thing lacking in Pollock's account is a justification of this way of striking the balance. He does argue in chapter twelve that we can normally expect that small plans will combine in consistent ways. There will be many exceptions though, and the hard question is whether there will be more unsatisfactory results from this compromise than from others. To put it differently, the book proceeds by developing the ways finite agents can calculate their best actions, in greater and greater scope, with each stage seeming plausible, until the scope gets large enough that the whole thing goes pop. Then the moral is supposed to be that since at every stage we were making good use of our resources, the picture that emerges, including the pop, is inevitable. But it may not be: a compound of maximally efficient procedures may not be maximally efficient.

One worry here is Pollock's implicit characterisation of the resources that his routines spare. They are supposed to go easy on memory and computing power, and thus in the human case, presumably, on long term memory, working memory, and time. This assumes that in carrying out a fairly intricate routine that happens to be reasonably economical, one will not make mistakes. But humans have the additional limitation that they slip, and it is an obvious possibility that one routine may be more efficient than another at the price of being harder to follow accurately. Some of Pollock's procedures certainly look as if I would have trouble doing them right. In fact I would suspect

that there is a tendency for routines that minimize computational load also to increase the likelihood of mistakes. There is a deep ambiguity here. We may take our aim to be to describe ways that if followed would solve our problems at a reasonable price. And we may take it to be to describe procedures that we can profitably try to follow. Both are valuable aims, but they are certainly not the same. Pollock's aim seems to be the first, though he helps himself to some of the normative vocabulary that is more often associated with the second.

Department of Philosophy
University of Alberta
Humanities Centre 4–115
Edmonton, AB, T6G 2E5
Canada

ADAM MORTON