May 15th, 9:00 AM - May 17th, 5:00 PM

Commentary on Woods

Robert C. Pinto

COMMENT ON JOHN WOODS,
"THE NORMATIVE IMPOTENCE OF IDEAL MODELS"

Robert C. Pinto
Department of Philosophy
University of Windsor
©1998, Robert C. Pinto

1. I take it that by an ideal model theory Woods means a theory which purports to describe the behavior of perfect or ideal agents under perfect or ideal conditions. Roughly speaking, an ideal model theory of rationality will tell us what perfectly rational investigators or decision-makers will do under various circumstances—that is to say, how they will make up their minds about what to do or about what is the case in the face of various sorts of evidence, constraints, as well as epistemic and nonepistemic goals. An ideal model theory of economic behavior will tell us how rational agents in a market-place will behave given certain utility functions and certain conditions of scarcity or abundance. And an ideal model theory of argumentation will tell us how perfectly rational interlocutors will conduct themselves in the course of a "critical discussion"—perhaps describing the stages in which their discussion will proceed, the kinds of issues they will try to settle and in what order, the kinds of speech-acts they will engage in at various stages of their discussion, the conditions under which they will accept or resist each other's claims, and so on.

Such models can be used as norms for judging the behavior of actual investigators, decision makers, economic agents, critical discussants, and so on. An actual person's behavior will be pronounced rational (or good) precisely to the extent that it coincides with or resembles the behavior of a perfectly rational (or perfectly good) person, as outlined in the model. "Rational" and "good" are terms of appraisal,[1] and a model functions as a norm when it is used as the basis of an appraisal. I assume in what follows that a normative model is simply a model which functions as a norm.

It is an interesting question what the strengths and weaknesses of articulating norms in *this* fashion are, as opposed to formulating them in some other fashion. For one thing, where the model describes the behavior of ideal agents under ideal conditions (e.g., agents who possess perfect information), it may not be clear how to evaluate the behavior of agents whose actual circumstances are far from ideal. For another thing, where the terms employed to characterize the behavior of model agents are classificatory, rather than comparative or quantitative, there may be no easy way to view the behavior of actual agents as approximating to or resembling that of model agents. The problems faced, and the solutions available, will vary from domain to domain, and therefore the prospects for utilizable ideal models will have to be judged on a domain by domain basis.

Though Woods is clearly aware of these issues, they aren't the one he wants to press. Rather, Woods' question is about the grounds or justification for adopting one, rather than another, ideal model as a *norm* on the basis of which to assess the behavior of actual agents. It is in fact always possible to formulate a variety of models which *could* be made the basis for assessing actual agents or actual performances.[2] And therefore anyone who chooses to employ this model rather than that one can always be asked her or his reasons for choosing it rather than an alternative as the basis for appraisal.

As I see it (and I doubt that Woods sees it any differently) this kind of question arises with respect to any sort of

norm—not just with respect to ideal models when they function as norms. If the norms which guide our appraisals be goals, then why these goals rather than some other? If the norms be rules or principles, then why these rules or these principles, rather than some other? Moreover, I'm not inclined to think that Woods has given us—or even that he intended to give us—reason to believe that those who employ ideal models as norms are any worse off than those who employ other sorts of norms. If the general problem is as intractable as Woods appears to suggest it is, then it is not just ideal model theories, not just normative theories generally, not even just philosophy, whose end can be announced. One wonders how any sort of appraisal can be other than arbitrary.

2. To see whether my alarmism is warranted, we need to look more closely at what Woods calls the bootstrapping problem.

Woods' formulation of that problem draws on features that are specific to normative ideal models. Ideal models of the sort we're interested in describe the behavior of ideal agents under ideal conditions, and therefore can be construed as containing "laws" which govern the behavior of such agents. Those "laws" typically—perhaps necessarily—differ from the laws that govern or describe the behavior of actual agents; they are, in Woods phrase, "empirically false." That is to say, typically the behavior of some actual agents differs from the behavior of "similarly situated" model agents. Indeed, without such shortfall, the model could hardly serve the purposes of appraisal—at least so long as the purposes of appraisal require the real world possibility of a negative verdict. With that in mind, we can understand why Woods describes the bootstrapping problem as follows:

> ... how is the [...] theorist to pick his "laws" from sets of empirically false generalizations,
> short of already having at hand a validated ideal model in which to rescue him from that
> very embarrassment? How, in turn, is he to pick a model as validated, short of stocking it
> with empirically false generalizations which, even so, have the property of "deserving", there, to be
> "laws" against which rational agency is measured? This is a difficulty serious enough to have a name.
> Let us dub it the *bootstrapping problem*.

To paraphrase: if the theorist doesn't already have a validated ideal model, he won't know which empirically false generalizations to pick as laws; but if he doesn't already know which empirically false generalizations to pick as laws, he won't be able to construct or validate an ideal model. He can't perform either task unless he's already performed the other; hence he's in business only if he can pull himself up by his own bootstraps.

Now this way of formulating the problem involves a highly questionable assumption that makes the problem seem quite thoroughly intractable. To wit, the assumption that our theorist can have a basis for "picking his laws" only if he has a validated ideal model at hand. Why should we grant that assumption? Clearly, we can sensibly expect anyone who would employ a particular model as a norm to have reasons for choosing that model as a norm. But just as clearly, it seems to me, there's no need that the reason for adopting a model as a norm involve appeal to an already validated model. There is a problem here, in the sense that there is a legitimate question about the bases on which we choose our norms. But it's not the problem that Woods wants to baptize the bootstrapping problem.

In a delightful bit of naughtiness, Woods intimates that some ideal model theorists might attempt to avoid the problem by the stratagem of stipulation: define the model by reference to the laws you want to pick; note that this renders the laws "analytic in the model" and commits the fallacy of *secundum quid* by treating what's analytic in the model as analytic full stop. He even goes so far as to quote a sentence from our pragma-dialectical friends, in which they suggest that the rules of their model "constitute a theoretical definition of a critical discussion." The naughtiness here consists in the insinuation that our friends do or would appeal to *that* claim about their rules as a

reason for appraising actual discussions in light of their model of a critical discussion. My skepticism about that insinuation stems in part from having heard these very friends offer quite different sorts of reasons and justification when the soundness of their norms were challenged—reasons that appealed to claims about the goals of critical discussion and the utility of their rules for achieving those goals.

3. Of course there are important questions hereabouts. And of course there was a time (in my mind a not very recent past) when it was fashionable to make it out that there are conceptual truths about how it is rational to proceed. For example, that expecting the future to resemble the past or engaging in enumerative induction is just "part of what we mean" by being rational. Indeed, what Woods read to us today is excerpted from a larger work in which, among other things, he takes on the notion of conceptual truth in what I think are original and illuminating ways. Yet even apart from more probing and detailed critiques, I'm quite ready to agree that appeals to conceptual truth, or to definitions of rationality, critical discussion, whatever—only postpone the difficult questions, and do not postpone them for very long. Let's grant, for the sake of argument, that by the very meaning of "rational" behaving in such and such a way is not rational. But then why is failure to be rational a *fault*, and why is conformity to the standards of rationality a *virtue*?3 That is to say, if one fleshes out "rationality" with one or another model or set of rules, it is still a matter of importance what reason there is for employing that model or set of rules as a norm on the basis of which to condemn or approve what people do and think. About this I think that Woods is absolutely right.

4. What Woods has to say about ideal models is part of a larger project in which he is criticizing—in other circles I might have said deconstructing—a widely held picture about the nature and justification of normative theories. Very roughly, normative theories consist of nomic generalizations which originate in and extend the data supplied by "intuitions," while at the same time those theories are able to correct the intuitions from which they've arisen and, perhaps most importantly, transform the untutored judgements called intuitions into principled judgements warranted by the theories to which the intuitions gave rise. Let me call this picture the theory theory. If I may caricature and grossly oversimplify Woods' criticism of the theory theory, he argues that normative theories owe whatever warrant they have to the intuitions which prompt them, but intuitions can't bestow warrant on the theories because intuitions aren't knowledge, but belief.

I think that something of that broader project and argument motivates two of the things Woods said in the portion he read to us today. First, and I take this to be offered as a consequence of the idea that the bootstrap problem can't be solved:

> ... a theorist's norms are nothing but what upon reflection he thinks people should do.

Next, and partly as a consequence of this,

> For what is a philosophical theory about rational performance if not a theory that tells
> us in a principled way what people should and should not do? On the present suggestion,
> there are no such theories; hence there are no philosophical accounts that tell us in a principled way
> what people should and should not do. Philosophy, i.e., normative philosophy, is out of business.

My reaction to Woods' skeptical conclusion is that it's a consequence of the fact that his rejection of the theory theory is only partial. Woods wants to say: a theorist's norms are just what on reflection he think people should do—*and that's not good enough*, that doesn't amount to a theory!

Why not? If the theorist's norms are what he thinks upon reflection—after the consideration of the views of

others, after correction in light of difficulties and inconsistencies, after having attempted to act and live on their basis and assessment of the results of so doing—why aren't they good enough? Well, though such norms may say what people should do and should not do, they don't say it in a principled way. And why not in a principled way? It looks to me as though the reason is that they aren't supported in the way that the theory theory says they should be. But if the theory theory is misguided, why should *that* be a reason to withhold the honorific title of theory?

Notes

1. I follow Dennett, in wanting to use "rational" as "a general-purpose term of cognitive approval." See D. C. Dennett, *The Intentional Stance*, Cambridge, MA: MIT, 1987), p. 97.

2. I leave aside Woods' distinction between case-one and case-two normative models. As far as I can see, that distinction either disappears or ceases to be of interest for the very reasons Woods gives: normative models for human agents can't be spun off simply from either first-order logic or probability theory.

3. Recall Wesley Salmon's trenchant rejoinders to conceptualist parries. See for example W. C. Salmon, "Should we attempt to justify induction?" *Philosophical Studies*, vol. 8, 1957.