

*Philosophical Books*, 34 No 3, 1993, 177-9

Contractarianism and rational Choice  
Edited by PETER VALLENTYNE  
Cambridge, 1991. xiii +339 pp.

This is a collection of essays on David Gauthier's *Morals by Agreement*. The essays are concerned with the general project of founding an ethical theory on a hypothetical social contract, where both the hypothetical bargaining and the rationale for actual people sticking to its result are worked out in game theoretical terms. As several contributors explain, Gauthier's way of doing this is definitely not the only one. But the interest and importance of that general project is something that emerges very clearly from the book. It certainly was much clearer to me after reading the book. I now understand why Gauthier has got so much attention, in spite of all the hesitations it is natural to have about some of his central ideas, such as that of constrained maximization. Moreover much other recent moral and political theory can be seen as part of the same project. It's a big perspective.

The book begins with an exposition of Gauthier by Peter Vallentyne, followed by a largely new essay 'Why contractarianism?' by Gauthier. These must be placed here in part to make it unnecessary for the contributors to do their own background exposition. But they evidently could not be persuaded of this and generally each fills in the background she needs, so that by the end of the book the reader is well and truly grounded in elementary Gauthiology. The level of exposition is generally very high, though. Several of the essays are moderately technical, in that they use a fair amount of game-theoretical apparatus. Nevertheless these - the contributions by Danielson, Gaertner & Klemisch-Ahlert, and Kraus & Coleman -

are among the clearest and most helpful things in the book. This particular topic seems to require either mathematics or perfectly transparent language: the direct plain speaking of Geoffrey Sayre-McCord and Holly Smith's contributions also clearly pays off. I won't discuss each essay separately. But none are duds; each one is a contribution in its own right.

Gauthier's essay at the beginning of the book is more explicit than he has been before that what he is proposing is a revision or substitute for morality. It is something a little like morality which we would be rational to accept. So moral intuitions as such do not have much force. That is just as well, because several contributors, David Braybrooke, Jean Hampton, and Jan Narveson in particular, point out worrying tensions between the conduct that Gauthierian morality might allow and what we would normally think of as decent behavior.

The source of the tension traces generally to Gauthier's construal of the basis point for bargaining, especially in terms of the 'Lockean proviso'. (see Peter Danielson's first essay in particular) and to his description of what would constitute a reasonable bargain at that point. (These are different. What is not quite clear yet is which troublesome moral intuitions can be linked to differences about just one rather than the other of them.) Gauthier's aims make him a peculiar sort of realist about the moral. He takes it that there are right and wrong answers about morality as construed along his lines. And in his essay at the end of the book he definitely takes disagreements with his critics about rationality as crucial: either he is wrong or they are. But there are not such definite answers to disagreements about the moral as target of our culture's intuitions. (The position is rather like a mild eliminative materialism in the philosophy of mind.) It is

not clear to me which disagreements with his critics can be shrugged off by this limited moral/shmoral realism.

No contributor really takes full account of this revisionist quality of Gauthier's thought. One basic question that is thus not taken up is: what range of rational substitutes for morality is there?. Forgetting about moral intuitions and what is obviously just or unjust, can someone who decides to live her life in accordance with the kinds of contract a rational self-interested person would make with her fellow agents have any assurance that there is only one such contract waiting to be formulated/drawn up?

The argument that there is only one such contract is that rationality is constrained maximization and constrained maximizers will go for only one kind of contract. That is not very convincing (see David Copp's piece), and it's not clear Gauthier sticks consistently to it. But in any case the reasonableness of constrained maximization is seriously undermined by several contributors. Danielson (his second piece) and Smith are particularly persuasive here. The most striking point they make concerns the worry that if I agree to cooperate with you iff you cooperate with me, and you do the same, then we have a circularity: each depends on the other. Danielson and Smith, in different ways, show that this is a serious worry. Danielson makes a link with the Axelrod tradition to argue that constrained maximization will often not be the best-performing strategy. Under many conditions other strategies will do better. (But some of them, while society-supporting, seem to challenge our ideas of decent behavior even more than Gauthier's. Oh dear: where are we headed?) The importance of the conclusion is in part the challenge to uniqueness assumptions. What is most rational for one person

may depend, crudely, on what kind of people she is surrounded by and, more subtly, by the kind of person she is disposed to become. (There are interesting parallels here with issues of mutuality in the philosophy of language. Danielson unites Gauthier and Axelrod. Can someone unite Danielson and Habermas?)

I think the failure of uniqueness is the central point that emerges from the collection. Most of the contributors are impressed with the new contractarian point of view. Most of them are not convinced by what Gauthier has done with it. Some of them think that other conclusions are the right ones. But many of them think that the edifice is not nearly as monolithic as Gauthier claims. I have cited Smith and Danielson in this connection, but it applies to other contributors too. For example Christopher Morris argues that there are several ways within the general scheme of interpreting the status of non-contracting agents, such as non-humans, future generations, and vulnerable outsiders. The failure of uniqueness makes the revisionary nature of Gauthier's project especially significant. If there are several ways of filling in the details, is this because some other moral factor, rooted in something besides rational self interest, has to come in? Or is it because morality is not best reconstructed as a single set of constraints on rational agents?