

**Review of Bart Engelen's *Rationality and institutions: on the normative implications of rational choice theory*.**

**Saarbrücken: VDM Verlag, 2008. 276 pp.**

SHAUN P. HARGREAVES HEAP

*University of East Anglia*

The motivating thought behind this volume is that “if one wants to “make up the rules” of the game of life, one has to start from a realistic view of its players” (Engelen 2008, 2). This may seem self evident but it is an idea that rarely achieves the prominence which it deserves—not least in methodological debates. For example, much ink has been spilled around Friedman’s famous 1953 essay that brought a form of positivism/behaviourism to economics, but comparatively little around this point. Yet it is really rather important.

If a theory is to be used normatively to provide, say, policy advice, then it has to be realistic in important respects. Otherwise the theory cannot hope to provide advice about how to change *our* world: its advice will apply to some ‘other’ world (i.e., the one that is in the relevant respects captured by that theory). To be specific, in mainstream economics, guidance is usually based on the Pareto criteria: that is, a policy intervention is warranted when economic theory predicts that some people will become better-off through that intervention without making any worse-off. The judgement about what makes people better or worse off is quite precise for this purpose. It comes from mainstream’s theory of what makes people act (i.e., the assumption of individual rationality): people are better-off when they better satisfy their preferences. If it turns out that people are not rational in the sense of being instrumental preference satisfiers, then the theory can still be useful in predicting people’s actions (because they could act ‘as if’ so motivated) but the theory cannot tell us when and if they are better off. The latter depends on the theory of individual rationality being in important respects descriptively accurate on the matter of how we act and how such action affects our well being.

This is why Bart Engelen is right to begin his book with a discussion of individual rationality in economics.

One might think that if a theory failed this realism test (and so was providing advice for some other world and not our own), then that would prove a terrible handicap. There are two defences that are sometimes deployed to avoid this conclusion. The first, potentially respectable one, is to argue that the theory is providing advice on what should be done if we are to live up to whatever alternative (or ideal) version of life is encoded in the theory. I say respectable in the sense that the theory would still be doing some work in guiding us with respect to what to do even if only in a sort of hortatory way. Nevertheless it would, if this was the case, still require some supplement in the form of a set of arguments around what made this alternative world ideal. For some of the reasons that are related to what Engelen sets out in his discussion of the rational choice model of preference satisfaction, it is difficult to come up with convincing buttresses of this kind (I say more on this below when discussing the difficulties of evaluating the activity of preference satisfaction if that is all one does). The second defence is that theories need *not* aspire to guide: it is enough that they predict (or explain) what happens in the world. Personally, I am unpersuaded by this as I cannot imagine how knowledge of the social world could ever be separated from acting in it.

This, in turn, is why Engelen's arguments, in the first part of this book, are also important, because he finds the economic rational choice model to be descriptively wanting and this is a problem for the reasons I have just sketched.

In the second part of the book, Engelen gives a tour of the dominant economic instrumental conception of rationality in chapter 2, and contrasts this with an alternative expressive notion of rationality in chapter 3. The economic model for this purpose is characterised as maximising in the sense that one satisfies best one's preferences and this depends in the usual way on having a coherent set of preferences (i.e., so that it is meaningful to talk about satisfying them best); in addition, these preferences are taken to be exogenous and egotistical.

The key difficulties with this model come from what behavioural economics tells us about individual behaviour. In particular, there are the various anomalies with respect to belief formation and inconsistencies in the notion of a preference that have been identified in the laboratory, and there is copious evidence that people have 'other regarding' or 'process' preferences. Here, less is made of this first group of findings than of the second. In particular, Engelen focuses on whether

these non-egotistical preferences can be accommodated by the model. In his view, they cannot and this is largely because of the difficulty that, to put the point compactly, the self has in being selfless. That the glow from being selfless has to be unintended is another way of expressing this; or, to make a bridge to what comes in the next chapter, actions acquire motivating force when they are other regarding because they mean something and not because they have consequences.

This is a line of argument that I have also tried to make (and so I can't help but be sympathetic). Nevertheless, while I think the argument is based on a genuine distinction between types of motives to action that is important, I have become less persuaded that the concept of a preference is not sufficiently elastic to accommodate other-regarding or process-oriented actions (or alternatively that the concept of a consequence cannot be expanded to include the meaning of an action). This elasticity comes at a cost, however, and this is where the genuine distinction between the types of motive resurfaces, albeit in the language of preferences. Preferences acquire a two tier structure and they can no longer be taken to be exogenous as they are socially embedded and activated. Psychologists have yet another way of developing the same insight which has filtered through to some parts of economics: there are two types of reason, 'extrinsic' and 'intrinsic', and what is interesting is the dynamic between the two as when 'intrinsic' reason is 'crowded-in' or 'crowded-out'. In much the same way, Engelen concludes the chapter on expressive reason by arguing that the two types of reason are complementary: one cannot be reduced to the other, and the important task of social science is to decide when and where which type of reason is guiding action.

The next chapter gives an illustration of how some of these ideas can be set to work in unravelling the paradox of voting (chapter 4). There are two parts to this paradox. The first is why people vote when they can have virtually no effect on the outcome. This is sometimes answered by arguing that people find voting pleasurable but if this was the only reason, then there would still be no explanation of why people vote for the particular person that they do. This is the second part of the paradox. Engelen argues that people vote because it is expressively rational. Voting offers the opportunity to say something about oneself precisely because it cannot be construed as an instrumentally rational action, and it is because one is saying something about oneself that one can explain the choice of who one votes for. 'Voting is like cheering' is

the helpful way this is put: it is done largely to express support and there is no real expectation that bleating in this way, for example from the stands in a football match, will have any effect on the outcome.

This is persuasive and it is nicely argued. It also seems to me to be generalisable. The general point is that it is precisely when instrumental reason cannot give clear guidance to action that there is scope to express something through action. Otherwise the meaning of an action would be obscure even if one intended to express something, say moral, through action because it could equally be construed as selfishly instrumental. Voting fits the bill well because there is apparently no clear selfishly instrumentally rational interpretation. Some care is required here, though. It is not that it makes *no* sense to vote in all circumstances because, if nobody else votes, then one would influence the outcome by voting. Formally, the paradox is really bound up with the fact that this is a game where there are no *pure* strategy Nash equilibria. There is, of course, a mixed strategy Nash equilibrium, although I have never seen any attempt to explain voting by appealing to this solution concept even though it would formally resolve the so-called paradox. This must be because mixed strategy equilibria are so implausible that they could not be generally taken as supplying a reason for instrumentally rational agents to vote. The clearer class of games where instrumental reason fails to guide (and this is the point about generalisation) are games with multiple Nash equilibria and these would, therefore, be the circumstances where there is scope to instantiate (unambiguously) the norms which enable action to become symbolic.

The last part of the book turns explicitly to how the different conceptions of rationality construe the institutions of the market, state and communities. Buchanan's constitutionalism is used as the exemplar of what happens when you build institutions around the economic model of rationality. Thus chapter 5 provides a quick sketch of a set of familiar theses that: a) connect freedom with efficiency, b) turn Rawlsian and other collective choices into a choice over rules rather than outcomes, c) make unanimity in these matters all important, and which d) are alert to the drift to a wasteful form of big government through rent seeking and the like.

What seems so obviously wrong with this account is the way that politics becomes no more than the pursuit of self interest by another name. What has been lost is any sense that the political (and other non-market) arenas are spaces where ideas and argument (about how to live

and how to organise society) are tested and agreements are reached; and that having an institutional space where this goes on is very important.

What makes such a public space important is the way that, by their nature, arguments in these arenas have to be impartial because arguments can never be persuasive if they appear self evidently self serving. This, in turn, connects with the alternative expressive conception of rationality. 'People are concerned with a sense of self-worth' is the way that I would put it. They reflect on what they do and they like to find their actions worthy. The standards for such reflections can, however, *never* be purely personal, otherwise they could be self-serving and the judgement of worthiness in such cases would lose its psychological edge. This is why communities, groups and the political institutions of collective decision making are so crucial: they potentially provide the standards that are external to any single individual and to do so they have to be governed by a different currency (e.g., intrinsic reason, if we shift to the language of psychology).

Let me put it differently. If you are solely a preference satisfier, how could you know that the pursuit of preference satisfaction was a worthy activity if you lived in a society where the political and other social non-market institutions were merely another set of arenas where individuals pursued their preferences? Of course one can naturalise the pursuit of preference satisfaction and so deny that there is an issue, but this would be to fly in the face of much of what we know psychologically about humans. What cannot be done is to appeal to a meta-preference that we have to act in this way. Preferences (including the 'meta' ones) are just that: they do not provide reasons for action. And if we do, indeed, care about whether we have good reasons for action, then we need a shared external space where those reasons are manifest (i.e., where they are debated, discussed, and tested) because purely private ones will not do the trick.

Chapter 6 makes these points systematically. Institutions should be designed in the knowledge of how people behave: they should recognise the role of communities as the locus for judgements about intrinsic value, the role of the state in mediating between communities, and the dependence of the market on norms of pro-sociality and trust which can all too easily be crowded-out. That is, the main challenge thus lies in "developing an institutional structure such that states, markets and communities are mutually enhancing" (Bowles and Gintis 2002, F431; quoted by Engelen 2008, 234).

This quote from Bowles and Gintis is what concludes this part of the book. I agree completely with it, but it also flags up an earlier discussion of Bowles and Gintis in this chapter that fits less comfortably with the rest of the book. Let me just sketch how Bowles and Gintis are used earlier in this chapter. They are important in Engelen's view because they supply an evolutionary account of pro-social behaviour within a group which depends on the existence of some competition between groups. This is why the conception of citizenship and the institution of the state are potentially important in the way sketched above: they are the mechanisms through which we escape the evolutionary legacy of group conflict without, if we are clever, losing the incentives to behave pro-socially that come from identification with groups.

I actually think this is an important and interesting argument which needs to be incorporated in the design of our institutions. Nevertheless, evolutionary arguments are not obviously helpful when accounting for why we are expressively rational in the sense that Engelen and I use the term. This is because evolutionary arguments can be constructed to explain the origin of pro-social *behaviours* but this is not the same as explaining why we come to attach symbolic significance to those behaviours. That is, it does not explain why we think such behaviours might be right, honourable, just, and so forth: this is the symbolic realm where expressive reason roams. I would not want to exclude the possibility of an evolutionary element in the explanation of this human faculty, but I cannot help but feel that the approach of current evolutionary arguments seems strangely to overlook the old methodological lessons that found behaviourism wanting. The analysis of behaviour is simply not enough.

This grumble at the end should not detract from the fact that this is a good book. It is exceptionally well written and its argument ranges across big literatures to draw important conclusions for the institutions of social life. What else can one ask for?

## REFERENCES

- Bowles, Samuel, and Herbert Gintis. 2002. Social capital and community governance. *Economic Journal*, Royal Economic Society, 112 (483): F419-F436.
- Engelen, Bart. 2008. *Rationality and institutions: on the normative implications of rational choice theory*. Saarbrücken: VDM Verlag.
- Friedman, Milton. 1953. The methodology of positive economics. In *Essays in positive economics*, Milton Friedman. Chicago: University of Chicago Press, 3-43.

**Shaun P. Hargreaves Heap** is professor of economics at the University of East Anglia, UK. His current research focuses through experiments on the social and emotional aspects of decision making and on the measurement of diversity in the media.

Contact e-mail: <s.hargreavesheap@uea.ac.uk>