City Research Online

City, University of London Institutional Repository

Citation: Denis, A. (1994). Rationality. In: Arestis, P. & Sawyer, M. C. (Eds.), The Elgar Companion to Radical Political Economy. . Cheltenham, UK: Edward Elgar Publishing. ISBN 184376864X

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/4049/

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

Rationality

Andy Denis

From its inception in the marginal revolution of the 1880s, neoclassical economics has depended on the notion of a rational economic agent, a *Homo economicus*. Equally, the notion of rationality has been the focus of criticism from those wishing to dispute one or another aspect of mainstream economic thought. Lagueux (2010) provides an excellent introduction to the literature and the issues. Giocoli (2003, 2005) provides a detailed and highly readable account of the emergence of the neoclassical rational individual agent, in particular showing the supersession of the older conception of rationality as maximisation by the newer version of rationality as consistency. In the older presentation of the matter, the agent was a recognisably human individual attempting to optimise subject to constraints. In the later view, all that is left is consistency: "The major meaning of rationality is a condition of consistency among choices made from different sets of alternatives" (Arrow, 1996, p. xiii). In this view, rationality is defined by the existence of a preference relation which is complete and transitive. Giocoli's point is that both maximisation and consistency are requirements of rational behaviour seen from different perspectives – failure to maximise would be inconsistent and inconsistency would imply a failure to maximise. So the displacement of one by the other has been a development, not of the 'body of science', but of the 'image of science' of economics – how economics presents itself both to itself and to the rest of the world.

Rational behaviour is behaviour in accordance with reason, behaviour which in some sense serves the actor's interests. Most writers seem to accept that rationality is an essential premiss for any science of economics. In so far as economics is a science of human action, there seems to be little one can say of action which is unreasonable. However, much work has gone into distinguishing alternative concepts of rationality which, it is suggested, could serve as this premiss (Hargreaves Heap, 1989). In particular, following Max Weber, economists have distinguished between a range of rationalities including instrumental, value, substantive, procedural and expressive rationality.

Lagueux(2010) has convincingly shown that some of these are not alternative concepts, but the same concepts seen from different angles. For example, substantive rationality describes decisions which do in fact promote the goals of the actor, while procedural rationality describes the process of arriving at those decisions. It is difficult to see how a process of decision making which leads to substantively non-rational behaviour can be regarded as procedurally rational, and vice versa. The opposite of (substantive) rationality is not some other kind of rationality but irrationality.

A similar approach might be applied to value rationality. It is held that it is instrumentally rational to behave in a way which promotes a particular goal, but the selection of the goal itself is an instance of value rationality: it is said to be more rational to select some goals than others, life over death, pleasure over pain, and so on. However, this appears to rest on a confusion. If goals are truly *ends*, then they are not selected by agents at all, but given to reason by the passions to which "reason is … slave" (Hume, 1888, p 415). On the other hand, if they *are* selected by agents, then they are

selected as *a means* – a means to an end lying beyond them, an end which is itself given to reason by the passions. This is not to say that there are not higher and lower goals, only that it is inappropriate to deem them the spheres of *different kinds* of rationality, and, in particular, that it is 'value-rational' to adopt some high-level goals and 'value-irrational' to adopt others.

A related argument may be made regarding the distinction between parametric and strategic rationality (Elster, 1984). It is certainly the case that rational agents will act differently, depending on whether they expect other agents to respond (strategic behaviour) or not (parametric behaviour). But to regard this as exemplifying two different *kinds* of rationality seems unjustified. Finally, the opposition between expressive and substantive rationality is vulnerable to the same criticism as that made of the instrumental-value polarity above. If one chooses, for example, to give blood, in order to express who one is, then we can equally describe one's choice of where to live, what occupation to follow, how much leisure to take, what form of transport to use, and indeed choices regarding every other aspect of one's life as instances of the same kind of rationality. The behaviours are once again simply adopted as instruments or means to ends which are given to reason: they exemplify substantive rationality.

A major consequence of the assumption of rationality is that the rational agent will optimise. That is, that, given all the constraints he faces, he will do what is best suited to promoting his interest. This is rational because, given that it is possible for him to take this action, any alternative action would be suboptimal, and therefore irrational. This simple point has been subjected to a firestorm of criticism, in particular from a tradition following writers such as Herbert Simon (1976), and Kahneman and Tversky (1979), and which has come to be known as behavioural economics. In particular, it is said that agents may not optimise, because

- they may lack the information to do so;
- they may lack the computational power to do so, and because;
- they may lack the motivation to do so, in that they may adopt unselfish goals.

It is, however, completely mistaken to see these points as in any way a challenge to the neoclassical paradigm, rather than a refinement and development of neoclassical thinking. It is entirely possible for neoclassical economists to adopt these points and correspondingly to develop the constraints the agent is working under, and the rational responses available. The agent with limited information and computational power, the 'boundedly rational' agent, will develop stratagems for coping including rules of thumb and social institutions which economise on information and computation. The criticism that agents may adopt unselfish goals is not to the point since it is not a requirement of neoclassical thought that the individual will only adopt the most narrowly selfish goals. On the contrary, the agent may place whatever weight he feels like on the welfare of other individuals or society as a whole, or such ideals as literacy and charity. The point is that, once adopted, they are arguments in *his* utility function, and are then treated as *his* interests. The criticisms of the 'selfishness' of neoclassical agents is largely as misplaced as the similar criticism of the 'selfishness' of genes in the evolutionary theory of Richard Dawkins (Dawkins, 1989, p. 267).

The problem with the neoclassical notion of rationality lies elsewhere. In particular, the neoclassical stance assumes that micro-level rationality, the rationality of individual agents, is translated by the market system into macro-level rationality, the desirability of the social consequences of the actions of rational individual agents. This belief allows neoclassical economics to argue that laissez-faire

works: rational individual behaviour underpins desirable social outcomes, so there is no need for society to intervene by means of its political organisations, its states. There is an alternative view, articulated by such writers as Keynes and Marx, (Keynes, 1936, p. xxxii; Marx, 1975, p 87) that social pathology *emerges* at the macro level, that individually rational behaviour may issue in collectively irrational consequences, if the incentive structure facing individuals is such as to guide them to undesirable outcomes. It is interesting to see *both* standpoints articulated in a recent and well-known article by Paul Krugman (2009) on the financial and economic crisis which emerged in the late 2000s. Alongside repeated statements that markets might not be perfect we see repeated condemnation of the notion of rational agents: "economists fell back in love with the old, idealized vision of an economy in which rational individuals interact in perfect markets, this time gussied up with fancy equations … They turned a blind eye to the limitations of human rationality that often lead to bubbles and busts" (p. MM36).

Krugman's (2009) approach thus binds together two distinct stances. One is to adopt the Keynesian view that markets do not spontaneously generate optimal outcomes, the other is to adopt the neoclassical view that markets automatically transform individual rationality into collective rationality, and hence, that if there is collective irrationality it must be due to individual irrationality. For the latter argument, behavioural economics and behavioural finance come at a very helpful moment. Similarly Akerlof and Shiller (2009) in a widely celebrated book on 'animal spirits' argue that "standard macroeconomists have [focused] on how the economy would behave if people had only economic motives and if they were also fully rational" (p. 168). From this perspective what is wrong with modern macroeconomics is that it fails to appreciate the breadth of motivations that individuals might have, and the fact that they might not act rationally – both intuitions which neoclassical economics has little difficulty impounding. From the contrasting perspective, with roots in the writings of Marx (1975, 1976) and Keynes (1936) amongst others, any such attempt to understand macroeconomic pathologies, such as unemployment and crises, by tweaking the rationality of the agent incorporated in the model, looks very much like trying to understand a traffic jam by looking under the bonnet of the individual vehicle (Hofstadter, 1985).

This raises the issue of substrate neutrality (Dennett, 1995). Traffic jams and unemployment, like Darwinian evolution in Dennett's account, are founded in substrate-neutral processes. A traffic jam is consistent with a range of different types of engine, from internal combustion engines to horse-drawn wagons, and even pedestrian traffic. Again, the same vehicles which are involved in a traffic jam in a busy town could be running smoothly on the motorway. Equally, unemployment and economic crises are compatible with a range of individual behaviours including both standard neoclassical rationality and bounded rationality. Equally, the same individuals in a different context could be participating in a well-functioning economy. Within limits the same micro-foundations, including individual agent rationality, are consistent with multiple outcomes, and the same outcome is compatible with a range of substrate properties. This is not to suggest that it is impossible for mechanical failure in an individual vehicle to cause a traffic jam, or for irrational behaviour to cause economic crises. What is suggested rather is that it is an empirical issue: each instance of macro-level suboptimality has to be investigated to see whether its roots lie in individual failure or in perverse consequences of the network of social relations within which each individual is embedded. The neoclassical reductionist assumption that the social world is a congeries, and that individual and social rationality are therefore directly linked, hobbles economic science, reducing it to apologetics.

Daniel Dennett (1971) identifies "the assumption of rationality" with the intentional stance. An intentional system is one whose behaviour can be understood by ascribing to it beliefs and desires. The intentional stance predicts behaviour "by ascribing to the system the possession of certain information and by supposing it to be directed by certain goals and then by working out the most reasonable or appropriate action on the basis of these ascriptions" (p. 90). Neoclassical economic agents are intentional systems in Dennett's sense. Following Dilthey and Weber (Gerth and Wright Mills, 1948, p. 56), this is sometimes expressed in terms of *Verstehen* – the interpretation of human action by means of the observer's human empathy with the actor. This has been taken up particularly within the Austrian School. For Hayek (1979), for example, in order to understand society we must start with "the concepts and views held by individuals which are directly known to us" and proceed from these elements "to build up the more complex phenomena" of society: "it is the attitudes of individuals which are the familiar elements and by the combination of which we try to reproduce the complex phenomena, the results of individual actions, which are much less known" (p. 65). It is this starting from the individual agent, with whose beliefs and motives we can empathise, which, for Hayek (op. cit.) , qualifies our method as methodologically individualist.

This, however, is only the beginning. We can understand the behaviour of agents because like us they are human, like us they have beliefs and motives, like us they are intentional systems: "we can recognize these elements of human relationships only because they are known to us from the workings of our own minds" (Hayek, 1979, p. 59). But the intentional stance can be applied to *any* system in which it makes sense to speak of information and goals. An obvious non-human example is that of genes. The selfish gene programme (Dennett, 1995; Dawkins, 1989) is an application of the intentional stance to DNA. For Hayek (1979), the subject matter of social science is the "network of relationships" within which individuals are embedded, "the social structure [which] can remain the same although different individuals succeed each other at particular points", "the constant structural element which can be separated and studied in isolation" (p. 59). Again, for Hayek (1978), this social structure, this network of relations, like DNA, has been through an evolutionary process. It remains a challenge for social science to apply "the assumption of rationality" and the intentional stance, as exemplified in the selfish gene programme, to the evolution of institutions. That is, to recognise that social forms survive and evolve, not in order to serve human purposes, but because they are good surviving, good at levering themselves into subsequent generations. This raises the prospect of supra-individual social entities enjoying their own rationality, and coordinating social activity in their own interests, interests different from or even opposed to those of their human substrate. States (Marx, 1975) and corporations (Marx, 1976) are immediate candidates for this kind of analysis.

References

Arrow, K. J. (1996) "Preface" in KJ Arrow, E Colombatto, M Perlman (eds) *The Rational Foundations of Economic Behaviour* London: Macmillan, xiii-xvii.

Akerlof, G.A., and Shiller, R.J. (2009) *Animal Spirits. How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism* Princeton and Oxford: Princeton University Press.

Dawkins, R. (1989) [1979] *The Selfish Gene* Oxford: Oxford University Press.

Dennett, D.C. (1971) "Intentional Systems" *The Journal of Philosophy* 68(4): 87-106.

Dennett, D.C. (1995) *Darwin's Dangerous Idea. Evolution and the Meanings of Life* London: Penguin.

Elster, J. (1985) *Ulysses and the Sirens. Studies in rationality and irrationality* Cambridge: Cambridge University Press.

Gerth, H.H., and Wright Mills, C. (eds) (1948) *From Max Weber. Essays in Sociology* London: Routledge & Kegan Paul.

Giocoli, N. (2003) *Modeling Rational Agents. From Interwar Economics to Early Modern Game Theory* Cheltenham: Edward Elgar.

Giocoli, N. (2005) "Modeling rational agents. The consistency view of rationality and the changing image of neoclassical economics" *Cahiers d'économie politique* 49(2): 177-208.

Hargreaves Heap, S. (1989) *Rationality in Economics* Oxford: Blackwell.

Hayek, F.H. (1978) *New Studies in Philosophy, Politics, Economics and the History of Ideas* London: Routledge & Kegan Paul.

Hayek, F.H. (1979) [1952] *The Counter-Revolution of Science. Studies on the Abuse of Reason* Indianapolis: LibertyPress.

Hofstadter, D.R. (1985) *Metamagical Themas. Questing for the essence of mind and pattern* New York: Basic Books.

Hume, D. (1888) [1739] *A Treatise of Human Nature: An Attempt to introduce the experimental Method of Reasoning into Moral Subjects* London: Oxford University Press.

Kahneman, D., and Tversky, A. (1979) "Prospect Theory: An Analysis of Decision under Risk" *Econometrica* 47 (2): 263–291.

Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money* London: Macmillan.

Krugman, P. (2009) "How Did Economists Get It So Wrong?" *New York Times* 2 September 2009, p MM36.

Lagueux, M. (2010) *Rationality and Explanation in Economics* London: Routledge.

Marx, K. (1975) [1843] "Critique of Hegel's Doctrine of the State". In Karl Marx (1975) *Early Writings* Harmondsworth: Penguin, 57-198.

Marx, K. (1976) *Capital. A Critique of Political Economy. Volume One* London: Penguin.