

**Review of Samir Okasha and Ken Binmore's *Evolution and rationality: decisions, co-operation, and strategic behaviour*.
Cambridge: Cambridge University Press, 2012, 281 pp.**

WILJAN VAN DEN BERGE
Tilburg University

There is a long history of sharing models between evolutionary biology and economics. Most famous is of course the introduction of game theory from economics into biology, which was then transformed into evolutionary game theory and re-imported into economics (Grüne-Yanoff 2011). Another area where both biology and economics have things to say is on cooperation. Both evolutionary biology and rational choice theory (RCT, hereafter) have problems accounting for cooperation within their standard frameworks. This has led to the development of a literature on cooperation that straddles the borders between economics and biology, and also philosophy.

These topics were explored in two conferences organized by Samir Okasha and Ken Binmore at the University of Bristol, and eleven papers from these conferences are collected in this volume. In the introduction, Okasha and Binmore emphasize the strong connections between evolutionary theory and RCT. Continuing a theme from Okasha's recent work (Okasha 2007; and 2011), they focus particularly on how both RCT and evolutionary theory conceptualize behaviour 'as if' it is maximizing some quantity: utility in the case of RCT and fitness in the case of evolutionary theory. They suggest that the main question arising from this is, "when is it possible to identify the economist's notion of utility with the biologist's notion of fitness?" (p. 2).

While this specific question might be of interest to some, it makes the book much more worthwhile that not all the chapters explicitly address it. Rather, the authors discuss a range of conceptual and methodological issues regarding the many interrelations between evolutionary theory and RCT, construed in their broadest senses. The most prominent interrelation is the use of game theory in both economics and biology to model strategic interaction. In addition, about half of the contributions address issues of cooperation in one way or another. Different explanations of cooperation, such as strong

reciprocity (Vromen, Gintis and Sterelny), team reasoning (Gold), or the evolution of preferences (Berninghaus, et al., and Wolpert and Jamison) are discussed. In this review I will try to give a flavour of the wide range of contributions in the book. I will divide them between the themes of *evolution and rationality* (broadly conceived), and *cooperation*.

I will begin with the two chapters on evolutionary game theory and then discuss some other ways of thinking about the relation between evolution and rationality. The first chapter, by Peter Hammerstein, provides a short introduction to game theory as it is used in biology that would be useful to those who are unfamiliar with the major developments in game theory. He discusses the main solution concepts and their potential usefulness in the social and biological sciences. Hammerstein argues that we should interpret players that play an equilibrium strategy (i.e., each choosing a strategy such that no player has an incentive to deviate from their chosen strategy) as resulting from learning that this strategy is most successful, rather than from strong rationality assumptions. He expresses his hope for a more explicitly Darwinian theory of learning.

In their chapter, Simon Huttegger and Kevin Zollman discuss the limits of what they call “ESS methodology”. An evolutionarily stable strategy (ESS) is one which, once generally adopted by a population, cannot be successfully invaded by any alternative strategy. Huttegger and Zollman criticise the standard method, within evolutionary biology, of only finding out whether a given strategy would be an ESS, but not considering the evolutionary dynamics that underlie the equilibrium. This can lead to mistaken conclusions. For example, it could mean that we only consider an equilibrium that is very fragile, or that is highly unlikely to be reached. They argue for considering the dynamics of the system, as epistemic game theory does in economics.

Another question that often returns is whether our evolved behaviour is in some sense ‘optimal’. In their chapter, Claire El Mouden, Maxwell Burton-Chellew, Andy Gardner, and Stuart West remind us that the biologist’s assumption of fitness maximizing behaviour is really only that, an assumption. In practice, we should at best expect to find adaptive behaviour. They discuss many reasons for this, such as developmental restrictions on organisms and evolutionary forces other than natural selection. The same holds true for the assumption of optimal behaviour by humans. They argue that many of the reasons why we do not find optimization in evolutionary biology also hold in

economics. In his chapter, Alasdair Houston looks at one particular violation of optimal ('rational') behaviour, the violation of transitivity. He argues that it can be given an evolutionary explanation, for example in terms of the evolution of behavioural rules.

Henry Brighton and Gerd Gigerenzer look at another aspect of rational behaviour: are we rational in "large worlds" as well as "small worlds"? Small world problems are ones whose characteristics can be perfectly known, such as a lottery. In contrast, large world problems require coping with inherent ignorance, which turns uncertainty into risk. Leonard Savage, who coined the distinction, famously said that it was "utterly ridiculous" to use his theory to try to understand large world problems (Savage 1954, 16). Brighton and Gigerenzer focus on inductive inference. They distinguish between the internal model that an agent holds and the external data-generating process. An agent's internal model can diverge from the real world on which it is supposed to be based due to stochasticity, underspecification, and misspecification. As decision problems acquire more 'large world' characteristics, these sources of error play a larger role and it becomes difficult to find the optimal strategy. We can, however, look at what simple rules have worked in the past and gradually improve those. This is the "relativist strategy" that Brighton and Gigerenzer advocate.

Siegfried Berninghaus, Werner Güth, and Hartmut Kliemt discuss a different way of understanding the relation between evolution and rational behaviour. They argue that we should distinguish between long and short-run time-scales. The long term time-scale is where preferences evolve (which they characterise as "push", since it does not involve goal-directed agency); the short-run scale is where we find rational behaviour ("pull"). Thus, the objective payoffs that determine evolutionary success are mapped onto the subjective payoffs that determine behaviour in the 'games' we play in everyday life. At the proximate level, we can have behaviour that is for example motivated by a social norm. But in the long run this will tend to produce higher objective payoffs for the individual.

I will now turn to the contributions on *cooperation*, starting with one that takes issue with Berninghaus, and co-authors' view. David Wolpert and Julian Jamison argue that indirect evolutionary models of the sort Berninghaus, et al. propose are too complex and can therefore only handle very simple situations. We should rather understand the long-term process as a learning process and the short-run process

as choosing a subjective utility function. This choice is analysed as adopting a “persona”, that we commit to and that will determine our actions in a given situation. Over time we learn which personas lead to the best objective results in given situations. This can explain short-run cooperation, where people adopt a cooperative persona in a particular context, by pointing to the long-run consequences of this behaviour.

Natalie Gold discusses another explanation for cooperative behaviour: team reasoning. The basic idea is that instead of asking “what should I do?” players will ask “what should we do?” and then play their part in the team strategy. How team reasoning exactly works depends on the assumptions we make about the payoffs for the team and how people identify with a team. Views diverge on this. Robert Sugden argues that people need explicit agreements, whereas Michael Bacharach argues that cues from the environment induce a so-called “we-frame” of thinking.

Gold examines the difference between the team reasoning account and the idea of other-regarding preferences. She finds that both have similar mechanisms—objective payoffs diverge from the subjective perception of the game—but she argues that team reasoning involves a more fundamental transformation of agency. She concludes with an unconvincing discussion of which explanation is better. One criterion she proposes is that we have a plausible evolutionary story for the team reasoning account. But it would be very easy to develop such a story for the other-regarding preferences account and then we are still left with deciding which story is better. It seems to me that carefully designed experiments are a more promising approach, since there we can at least partly control for the characteristics that induce team reasoning (framing or explicit agreements) and other-regarding preferences.

Another recently proposed explanation of cooperation is that humans are strong reciprocators. Jack Vromen tackles this hypothesis. In particular he addresses the confusions surrounding the concepts of evolutionary and psychological altruism and selfishness. For example, strong reciprocity theorists argue that ‘strong reciprocity’ (conditionally cooperating and punishing those who do not cooperate) is genuinely altruistic at both the evolutionary and psychological level, in contrast to ‘reciprocal altruism’, which is selfish on both levels. Vromen clears up these confusions by carefully distinguishing the different notions of altruism involved. He argues that evolutionary selfishness (lifetime fitness maximization) need not imply psychological selfishness.

But perhaps, Vromen argues, it does not really matter what causes strong reciprocity in terms of ultimate desires. What matters are the conditions under which it arises.

Herbert Gintis continues his argument for a synthesis of the behavioural sciences (Gintis 2009), resting on the pillars of “gene-culture coevolution, the sociopsychological theory of norms, game theory, the rational actor model, and complexity theory” (p. 213). Gintis tries to relate a lot of theories and concepts to each other to come up with a synthesis, because he believes that inconsistencies between the social sciences are a huge problem for progress. While one might wonder about a lot of the links that Gintis tries to establish, I am also doubtful of the general usefulness of this exercise. It does not necessarily strike me as a problem that, for example, sociology offers predictions that are inconsistent with those of economics.

Kim Sterelny provides an interesting take on the same sorts of issues, but from a pluralistic perspective. He argues that it is not a problem that models are inconsistent, since they each highlight different aspects of a given target system. Sterelny compares the rational actor with self-regarding preferences (‘homo economicus’) model with the strong reciprocity model and human behavioural ecology model. He argues that strong reciprocity models are a useful extension of the standard homo economicus model that retains its generality or tractability. However, he argues that human behavioural ecology, which explains human behaviour in terms of a combination of fitness maximizing population-level processes and individual adaptivity, is less applicable to modern societies, since the complexity of modern societies has led to a decoupling of reproductive fitness from “what we want”, our utility.

In conclusion, philosophers of economics might feel that this book focuses too much on evolutionary theory, but I believe that it tackles a lot of questions that are of interest for them as well. It also shows that there is a lot we do not yet understand. For example, some people argue that human behaviour can be understood as optimal in some sense, be it fitness-maximizing or ‘rational’, whereas others doubt this very much. The interpretation and explanation of economic experiments also remains a significant concern, both in terms of violations of the standard rationality axioms and the appearance of cooperative behaviour. How should we explain this behaviour? *Evolution and*

rationality provides strong evidence that biologists, economists and philosophers have a lot to gain from discussing these issues together.

REFERENCES

- Gintis, Herbert. 2009. *The bounds of reason: game theory and the unification of the behavioral sciences*. Princeton: Princeton University Press.
- Grüne-Yanoff, Till. 2011. Models as products of interdisciplinary exchange: evidence from evolutionary game theory. *Studies in History and Philosophy of Science Part A*, 42 (2): 386-397.
- Okasha, Samir. 2007. Rational choice, risk aversion, and evolution. *The Journal of Philosophy*, 104 (5): 217-235.
- Okasha, Samir. 2011. Optimal choice in the face of risk: decision theory meets evolution. *Philosophy of Science*, 78 (1): 83-104.
- Savage, Leonard J. 1954. *The foundations of statistics*. New York: Wiley.

Wiljan van den Berge graduated from the research master in philosophy and economics at the Erasmus Institute for Philosophy and Economics (EIPE) in September 2012 with a thesis on the use of evolutionary models in the social sciences. He is currently pursuing a Master's degree in economics at Tilburg University.

Contact e-mail: <wiljanvdb@gmail.com>