Biol Philos
DOI 10.1007/s10539-013-9392-0

The origins of human cooperation

Samuel Bowles and Herbert Gintis: A cooperative species. Princeton University Press, 2011, ISBN: 978-0-691-15125-0

Samir Okasha

Received: 19 June 2013 / Accepted: 21 June 2013

Abstract Bowles and Gintis argue that recent work in behavioural economics shows that humans have other-regarding preferences, i.e., are not purely self-interested. They seek to explain how these preferences may have evolved using a multi-level version of gene-culture coevolutionary theory. In this review essay I critically examine their main arguments.

Keywords Co-operation · Altruism · Multi-level selection · Gene-culture coevolution · Social preferences · Strong reciprocity

Scholars down the ages have held varying opinions about what (if anything) makes the human species unique. Candidate answers include language, rationality, culture and cognitive sophistication, among others. In recent years another candidate has received considerable attention from evolutionary-minded scientists, namely the human capacity to engage in extensive co-operative interactions with non-relatives, and to pursue joint objectives. Plausibly, this capacity underpins the complex forms of social organization found in human populations, with characteristic social institutions such as markets, firms, parliaments, nation states and armies.

Samuel Bowles and Herbert Gintis's new book, *A Cooperative Species*, takes as its starting point the uniquely human propensity to cooperate extensively with others who are not their kin, and in particular to engage in 'strong reciprocity' and 'altruistic punishment' (i.e., to reward co-operators and punish defectors in social exchanges, even when doing so is personally costly.) They endorse the view that has emerged from the recent behavioural economics literature, namely that humans have 'other-regarding' or social preferences, i.e., are not solely motivated by

S. Okasha (✉)
Department of Philosophy, University of Bristol, Bristol, UK
e-mail: Samir.Okasha@bristol.ac.uk

Published online: 03 July 2013

personal gain but genuinely care about the welfare of others, particularly fellow group members, and derive enjoyment from punishing free-riders. Such preferences, manifested in ultimatum games, public goods games and others like them, appear to be ubiquitous among humans (though are heavily culturally influenced); they are the proximate psychological cause of altruistic and cooperative behaviour in humans, Bowles and Gintis argue.

The main aim of *A Cooperative Species* is to explain how such social preferences could have evolved. The authors' explanation draws on two overarching theoretical ideas: multi-level selection theory and gene-culture coevolution theory. While both of these ideas will likely be familiar to students of the evolution of human behaviour, there is nonetheless considerable originality in Bowles and Gintis's treatment. Using a combination of analytical modelling and simulation, and drawing on techniques from both biology and economics, they construct rigorous models to show how social preferences could have evolved in the hominid lineage. Unlike much modelling work in this area, they actually calibrate the parameters of their models empirically, drawing on anthropological and biological work to obtain estimates of group size, levels of within-group relatedness, migration rates, and frequency of inter-group conflicts in the Pleistocene. The result is an impressive fusion of theory and empirical data which merits careful study by anyone interested in human evolution.

Bowles and Gintis's main hypothesis is one of gene-culture evolution. The genetically-based tendency to co-operate with others coevolved with group-level social institutions such as food sharing, information sharing, consensual decision making, group defence and punishment of free-riders, they argue. These social institutions evolved culturally rather than genetically; but crucially, they influenced genetic evolution by creating an environment in which genetically-based pro-social behaviours could evolve. This is because the social institutions in question all serve to dampen the variance in reproductive success within groups, thus mitigating the strength of within-group selection against altruism and allowing between-group selection to dominate. The social institutions were needed in order for individual (genetic) altruism to spread; and conversely the presence of altruistic individuals in the population was necessary for the social institutions to evolve culturally, they argue.

This is an interesting and plausible theory, backed up by detailed theoretical models and supported by empirical data. However some readers may wonder why Bowles and Gintis apportion responsibility between genes and culture in exactly the way they do, and whether it is empirically justified. Their view seems to be that since behavioural economists' experiments have shown that social preferences are ubiquitous, they likely have a genetic basis. But as they themselves admit, these experiments have also found wide cultural variation in how people behave in experimental games, and a dependence of behaviour on a host of seemingly trivial and irrelevant factors. One *might* interpret this by positing a universal underlying genetic disposition to co-operate that manifests itself differently in different environments, but this interpretation does not seem mandatory. Alternatively, it seems conceivable that social preferences may simply have evolved by cultural rather than genetic evolution. Why Bowles and Gintis prefer a hypothesis of gene-culture evolution, rather than just cultural evolution, is not made fully clear.

Multi-level selection is a key component in Bowles and Gintis's theory. They adopt the orthodox 'Price equation' approach to selection in group-structured populations, according to which the overall change in the frequency of a gene depends on the relative magnitudes of within and between-group selection. If the gene codes for an altruistic behaviour, it will be disfavoured by within-group selection but favoured by between-group selection. Bowles and Gintis argue that multi-level selection is necessary to explain the evolution of human cooperation; mechanisms based on inclusive fitness and reciprocation of benefits will not do the trick. The former cannot explain cooperation among non-kin, they argue, while the latter depends crucially on the availability of public information and so is only plausible in small groups.

A central plank in Bowles and Gintis's case is that traditional economic theory does not have the resources to explain the extent and nature of human cooperation. A venerable tradition in economics invokes the 'folk theorem' of repeated game theory to argue that purely self-regarding agents may in certain circumstances be able to achieve cooperative solutions, if they are rational. The folk theorem teaches us that in an indefinitely repeated Prisoner's dilemma, for example, in which players receive public signals concerning the past moves of their opponent, and can condition their choice of action on their opponent's past choices, cooperation can be a self-regarding best response for both players. Thus cooperative outcomes become explicable, this argument suggests, *without* having to posit other-regarding social preferences and thus without having to worry about how they evolved.

Repeated game theory has led to some ingenious work, but Bowles and Gintis devote a whole chapter to demolishing this entire approach. They do not question the formal correctness of the folk theorem but rather its relevance. The theorem only demonstrates the existence of cooperative Nash equilibria; but this says nothing about how or why agents will arrive at them. This problem is compounded if the Nash equilibria in question are mixed, as they often are, for such equilibria cannot be strict; thus no individual actually has an incentive to play their equilibrium strategy. Finally, if players have private rather than public information about each others' past moves, as seems quite likely, then the folk theorem does not go through. Bowles and Gintis offer a pessimistic assessment of the ability of traditional economic theory to explain the phenomenon of human cooperation that interests them. Traditional game theorists will doubtless disagree; however Bowles and Gintis's authoritative discussion of the limitations of repeated game theory provides an excellent motivation for their evolutionary approach, and serves a useful function.

For those who have followed the recent round of debates over multi-level selection and inclusive fitness among biologists, Bowles and Gintis's position is an interesting one. They point out, rightly, that all mechanisms for the evolution of altruism rely ultimately on positive assortment, i.e., like associating with like, which can be achieved in many ways. In this they follow the orthodoxy among contemporary social evolutionists. However, they regard inclusive fitness and multi-level selection as substantively different hypotheses, applicable in different circumstances, and are thus (implicitly) rejecting the widespread view that the two are ‘equivalent’, i.e., that the choice between them is a matter of perspective rather than empirical fact. This equivalence thesis was hinted at by Hamilton in the 1970s, and has gained widespread acceptance in recent evolutionary discussions (e.g., Marshall 2011; Lehmann et al. 2007; Frank 2013). In my view Bowles and Gintis are actually quite right to keep multi-level selection and inclusive fitness separate. In forthcoming work I argue that they are only ‘equivalent’ in a weak predictive sense, in that allele frequency change can always be expressed in terms of either, but not in the sense of offering identical causal explanations for the spread of a pro-social allele (Okasha (forthcoming)).

Bowles and Gintis are well aware of the reasons which have led many biologists to be wary of multi-level selection, and defend only a limited, carefully circumscribed version of the idea. They endorse the argument, common to many cultural evolutionists, that multi-level selection is likely to have been particularly significant in the human species, for a number of reasons. First, our hominid ancestors lived in tribes or groups which frequently came into often lethal conflict; second, the potential gains for a group in which co-operative tendencies were present were likely large; and third, culturally-transmitted social institutions, such as food sharing and other forms of ‘reproductive levelling’, serve to dampen within-group selection and prevent between-group variation from being diluted by migration. Thus the human capacity for cultural transmission, combined with the population structure of early hominids, provided the ideal environment for multi-level selection to favour the evolution of individually costly traits.

The various components of this package of ideas have been defended before, but Bowles and Gintis combine them in a particularly convincing and systematic way, based on a model in which groups undergo selective extinction, with a group’s probability of success in a conflict dependent on the frequency of altruists in the group. They show that the existence of culturally-transmitted social institutions greatly expands the parameter space in which the between-group component of selection can dominate the within-group component, thus facilitating the spread of pro-social behaviours and tendencies. They then argue that empirically, the social and physical environment of the late Pleistocene may well fall within the parameter space in which the model produces the desired result.

Bowles and Gintis do not duck the hard questions. In many discussions of multi-level selection, the balance between the within and between-group variance is exogenous, or determined by factors outside the model; however Bowles and Gintis endogenize it by explicitly modelling the cultural evolutionary process on which (in their model) it depends. This process involves the spread of group-level social institutions which contribute to reproductive levelling. But how do such institutions get going in the first place? Bowles and Gintis argue that the individual behaviours which sustain them, e.g., food sharing, can arise without altruistic preferences, by direct or indirect reciprocity within small groups, in which high-quality public information may be available. (Recall that reciprocal altruism is not ‘real’ altruism, as the individual gains in expected payoff over their lifetime.) This then gives rise to a social norm to share food, which can be extended to larger groups, they propose (p. 180). This explanation may seem slightly ad hoc, and comes as something of a surprise given their previous dismissal of reciprocal altruism, but they are at least alive to the potential weak points of their theory.

Another place where Bowles and Gintis tackle head on an issue that others have ducked is in their discussion of cultural transmission. They point out that cultural transmission arises from our capacity to internalize norms of behaviour, a capacity that has allowed the spread of behaviours that are clearly fitness-reducing, e.g., smoking, sky-diving and contraception. This is of course a familiar point in the literature on memes. However they then go a step further and ask how natural selection could ever have produced the capacity to internalize norms in the first place, if the result is that people go on to do things that harm their fitness? The answer, they propose, is that some norms are fitness-enhancing, so the capacity to internalize them is an advantage; and other fitness-reducing norms can then hitchhike a ride. Here as elsewhere, Bowles and Gintis do not rest content with a verbal explanation but construct an analytical model to show that the envisaged evolutionary process really can work.

A characteristic feature of Bowles and Gintis's book is the attention paid to both proximate and ultimate questions. In effect, what they have done is to marry the behavioural economics literature, which has a predominantly proximate orientation, with the evolution of cooperation literature, which focuses mostly on ultimate explanations. It is natural to try to weave these two into a single narrative, as they do. Bowles and Gintis argue that different evolutionary explanations for how cooperative behaviour spreads will likely give rise to different proximate preferences. Thus if kin selection is the dominant evolutionary explanation we should expect humans to care about their relatives, while if reciprocation is the explanation then we should expect humans to condition their other-regarding behaviours on the behaviour of their social partners, and so-on. Bowles and Gintis argue, quite reasonably, that given the diversity of proximate motives actually found among humans, a variety of different evolutionary mechanisms are probably implicated. But the multi-level version of gene-culture coevolution plays an especially important role, they think.

Combining attention to the proximate and the ultimate is welcome, but a doubt remains about whether we can move from one to the other as easily as Bowles and Gintis assume. In particular, the implications that an evolutionary explanation has for the proximate mechanisms we should see, i.e., preferences, depends on what one assumes about humans' cognitive capacities. This point is illustrated in a seminal (and unjustly neglected) paper of Samuleson and Swinkels (2006), who ask why humans seem to attach utility to 'intermediate' goods, such as food, shelter and sex. A naive explanation, commonly given by evolutionary psychologists, is that since such goods correlate with reproductive fitness, evolution has programmed us to want them. But as Samuelson and Swinkels observe, this is a weak argument. For why did evolution not make us care only about fitness itself? The answer they give is that, because of our cognitive limitations, we are unable to accurately assess the fitness consequences of many of our actions; if we were able to, there would be no point in caring about intermediate goods for their own sake. Therefore, our preferences (or utility functions) have the structure that they do in part because of our limited cognitive powers and/or the limited learning possibilities within a single lifetime. This subtle argument shows how difficult the inference from 'is evolutionarily beneficial' to 'is something that humans should want for its own sake' actually is.

To be fair to Gintis and Bowles, they actually make a point closely related to the Samuelson and Swinkels point in a different context. Discussing the evolutionary origins of the social emotions such as pain, guilt and shame, Bowles and Gintis argue that these emotions would not be needed if we possessed perfect information and had unlimited cognitive capacities, for then we would be able to directly calculate the optimal course of action in each circumstance. It is because we cannot do this, owing to our limited cognitive powers, that emotions earn their evolutionary keep they argue, inducing us to perform the 'correct' actions. This is an interesting and plausible argument but it applies more widely than to social emotions; a parallel argument could be applied to the evolution of social preferences generally. This suggests that Bowles and Gintis are somewhat too quick to move from ultimate to proximate in their discussion of social preferences. Further theoretical work on this issue would be welcome, as would the integration of the Bowles/Gintis theory with the extant literature in economics on the evolutionary foundation of preferences (e.g., Robson and Samuelson 2010).

A Cooperative Species is not a particularly easy read, in part because the analytical models are developed in the main text rather than in appendices or separate boxes. Inevitably this slows the reader down, making it difficult to 'skip the hard bits' and move on to the general discussion. But in a way this is no bad thing, as the models and simulations are central to Bowles and Gintis's argument, not merely illustrative of the ideas, so a reader who wishes to grasp their theory has no option but to work through them. Overall Bowles and Gintis have produced a fine and serious book, full of careful and detailed argumentation, and refreshingly free from the polemics that so often afflict discussion of these controversial issues.

References

Frank SA (2013) Natural selection. VII. History and interpretation of kin selection theory. J Evol Biol 26:1151–1184

Lehmann L, Keller L, West S, Roze D (2007) Group selection and kin selection: two concepts but one process. Proc Natl Acad Sci USA 104(16):6736–6739

Marshall J (2011) Group selection and kin selection: formally equivalent approaches. Trends Ecol Evol 26(7):325–332

Okasha S (forthcoming) On the 'equivalence' of inclusive fitness and multi-level selection: an approach using causal graphs

Robson AJ, Samuelson L (2010) The evolutionary foundation of preferences. In: Benhabib J, Bisin A, Jackson M (eds) Handbook of social economics, vol 1A. Elsevier, Amsterdam, pp 221–310

Samuleson L, Swinkels JM (2006) Information, evolution and utility. Theor Econ 1:119–142