

A Trojan Horse for Sociology? Preferences versus Evolution and Morality

Geoffrey M. Hodgson

Hertfordshire Business School, University of Hertfordshire, Hatfield, UK,
g.m.hodgson@herts.ac.uk

ABSTRACT

Herbert Gintis and Dirk Helbing have developed a highly impressive, over-arching theoretical framework, using rational choice theory, general equilibrium theory, and game theory. They extend this to cover “moral, social and other-regarding values,” plus social norms, culture, and institutions. While accepting the value of their contribution, I argue that there is a tension within their work between their depiction of the rational choice framework as a general “expression” of behavior and searching for *explanations* of, and detailed *motivations* for, particular phenomena such as punishment, altruism or moral sentiments. There is also a danger of over-generalization where a framework is stretched to cover every possible behavior. Indeed, rational choice theory with “other-regarding” preferences is strictly unfalsifiable. Furthermore, because “other regarding” agents are also depicted as maximizing *their own* utility, this framework cannot encompass adequate notions of altruism or morality. Instead, we should follow Darwin in seeking to explain the evolution of morality as a distinctly human motivation. The article concludes with a discussion of the implications of the Gintis–Helbing arguments for the future of sociology as a separate discipline.

Keywords: Preferences, Rational choice theory, Utility, Altruism, Morality, Evolution, Sociology, Economics

The paper by Herbert Gintis and Dirk Helbing – in which they offer an analytical core for sociological theory – is interesting, path-breaking, erudite, and elegant.¹ Their essay attempts to develop the analytical core of modern economics (including rational choice theory, general equilibrium theory, and game theory) to include “moral, social and other-regarding values,” plus social norms, culture, and institutions. Consequently, a framework taken from the core of modern economics is expanded to take on board issues that have traditionally been taken as the domain of sociology.

My comment concerns the nature of this project and its theoretical sufficiency. There is no doubting its theoretical value, but I question whether the (important and often valuable) search for theoretical generalization and encapsulation may distract attention from important historical, institutional and other specificities.

This complaint has been made by numerous economists in the past. Thomas Robert Malthus was concerned that David Ricardo’s theoretical system involved excessive simplification and generalization. Karl Marx insisted that the principal object of analysis should be capitalism as a historically specific mode of production, as did the German historical school, and the original American institutionalists (Hodgson, 2001). Other prominent economists from Alfred Marshall (1885) to Frank Knight (1933) took this point on board, while also developing a more general theoretical or conceptual framework to guide specific theories.

Generalization is often very useful. To some degree it is unavoidable, but it should not be taken so far as to sideline important specificities or to weaken the analytical or explanatory power of over-arching theory. In 1904, Max Weber (1949, pp. 72–80) argued that “the most general laws” are “the least valuable” because “the more comprehensive their scope” the more they “lead away” from the task of explaining the particular phenomenon in question. Over-generalization can mean losing explanatory power. Does the apparatus of rational choice theory, with other-regarding preferences, generalize a step too far?

In particular, I question whether the important human feature of morality can be adequately expressed within a rational choice framework. On this issue I concur with much of the critique of rational choice theory by Amartya Sen (1977, 1987a, 1987b) and some more recent arguments by Vernon Smith (2013). But unlike Sen, and like Gintis and Helbing, I wish to ground my arguments in evolutionary theory (Hodgson, 2013). In this respect, while addressing the evolution of morality, I find no better ally than Charles Darwin (1871).

I argue below that there is tension in the Gintis–Helbing paper between, on the one hand, the grand project of generalization via the rational choice

¹I wish especially to thank Herb Gintis and Vernon Smith for very helpful discussions.

framework and, on the other hand, their pointing to particular evolved dispositions, including “other-regarding values.” There is a contrast between using the copious rational choice framework merely to “express” manifest behaviors (in any living species), and identifying and providing an evolutionary explanation of distinctive human dispositions. Following Thorstein Veblen (Camic and Hodgson, 2011) and others, I suggest that the appropriate Darwinian and evolutionary route involves the latter, and much less the former: it involves explanation and discussion of motivation, rather than frameworks that merely “express” behavior.

Rational Choice Theory: Non-Explanatory, Non-Motivational, and Non-Intentional

Gintis, Helbing and their collaborators reject the self-regarding preference function that dominated mainstream economic theory from the 1870s to the 1990s. They point to the ample experimental and other evidence that contradicts this narrow and selfish version of rationality. They adopt a notion of “other-regarding” preferences that is broadly consistent with the empirical evidence.

They have also rebutted the claim that the empirical evidence falsifies the rational choice framework, or sustains instead a notion of “bounded” rationality. For them, the critics throw out the baby with the bathwater. Once the rational choice framework is expanded to include the possibility of other-regarding preferences, or of utility maximization that differs from payoff maximization, then it becomes consistent with the empirical evidence. I agree with this argument. My criticisms lie elsewhere.

In dealing with the debate over the rational choice model it is vital to distinguish between utility maximization and (expected) payoff maximization. A payoff is a known reward in a game that has *an explicit expected worth* (such as a declared monetary reward). Utility maximization is not necessarily payoff maximization, unless there is a monotonic relation between utilities and payoffs. This monotonic relationship is often assumed, but never in principle can be demonstrated, because utility itself – even if it is inferred from behavior – is unobservable. There is much evidence to refute universal payoff maximization, but – as Gintis, Helbing, and their collaborators have argued – this does not overturn the claim that individuals are behaving in a manner consistent with utility maximization. Gintis (2007, 2009) elsewhere considered much of the experimental evidence and pointed out that the absence of payoff maximization does not mean that players are behaving inconsistently or failing to maximize utility.²

²Note that in his earlier formulations, Gintis seemed more concerned to consider reasons and intentions. For example, Gintis (2007, pp. 11–12) brushed aside evidence that people are

Gintis and Helbing made a number of further important clarifications of their position. They wrote: “[t]he rational choice model *expresses* but does not *explain* individual preferences.” In other words, the rational choice model does not have explanatory power, at least in regard to explaining individual preferences. Furthermore, it is neither a theory nor a depiction of individual motivation. Hence they wrote: “the rationality assumption does not suggest that Alice is ‘trying’ to maximize utility or anything else.”

I do not disagree with this depiction of the rational choice model as one lacking in explanatory power and a theory of motivation. Other rational choice theorists concur with the Gintis-Helbing formulation. For example, Richard Posner (1980, p. 5) saw the “rationality of ‘economic man’” as “a matter of consequences, not states of mind.” Posner (p. 53 n.) continued: “in suggesting that people are economically rational, I am not making any statement about their conscious states. Rational behavior to an economist is a matter of consequences rather than intentions.” The differences between Posner, on the one hand, and the arguments of Gintis and Helbing, on the other, may be on the degree to which preferences are to be taken as “other-regarding,” but not on in their non-explanatory, non-motivational, and non-intentional characterization of rational choice theory.

As Ronald Coase (1977, p. 488) remarked: “[t]o say that people maximize utility tells us nothing about the purposes for which they engage in economic activity and leaves us without any insight into why people do what they do.” This is why the use of utility-maximization as a mere summary of behavior is inadequate.

Rational Choice Theory: Non-Falsifiable but not a Tautology

While abandoning payoff rationality, Gintis and Helbing defend a broader concept of rationality, involving consistency of behavior. Given behavioral consistency (or transitivity) – along with the other standard assumptions such as independence, continuity and completeness – it is possible to construct an ordinal utility function where behavior is consistent with expected utility maximization.

But there is a major difficulty. When the young Paul Samuelson (1937, p. 156) discussed utility maximization, he understood that “all types of observable behavior might conceivably result from such an assumption.” Because

not very good at solving logical problems with his questionable claim that “most individuals do not appear to have difficulty making and understanding logical arguments in everyday life.” But the evidence of Cosmides (1989), and Cosmides and Tooby (1994a, 1994b) suggests otherwise. Even from Gintis’ standpoint, his argument seems somewhat superfluous, because if an individual made the same logical errors over and over, then he or she might be behaving consistently, and might be “rational” by his criterion.

utility is unobservable, all kinds of behavior can be “expressed” in terms of utility maximization, without fear of refutation. As Sidney Winter (1964, pp. 309, 315), Lawrence Boland (1981) and others have argued, no evidence can possibly refute the theory that agents are maximizing some hidden or unknown variable (such as utility).

If experiments show that some consumers appear to prefer a monetary reward that is less than the expected outcome, or appear to have intransitive preference orderings, or defy the independence axiom, then we can always get round these problems, and make the evidence consistent with utility maximization, by introducing other explanatory variables.³

For example, assume that a subject is faced with a choice between \$10 with certainty, and \$1,000 with a probability of 2 per cent. Experiments with real subjects indicate that in such situations the \$10 option is sometimes chosen, despite the fact that the expected value of the second option is higher at \$20 (Slovic and Lichtenstein, 1983). But this does not falsify expected utility theory, once we accept that (expected) utility is not necessarily measured in terms of the monetary payoffs in the experiment. If we assume an added disutility associated with involvement in a risky and low-probability choice, then the theory that people are maximizing their utility is not overturned by these experiments. A risk-averse actor may not maximize expected monetary value but still be maximizing expected utility. By appropriate functional manipulation, the choice of \$10 can be made consistent with the maximization of expected utility.

Gintis and Helbing might respond that inconsistent behavior would refute utility maximization. The problem here is one of identifying inconsistent behavior in empirical terms. On repeated visits to the same restaurant, we may prefer steak to fish one day, and fish to steak on another. Is this behavior inconsistent? Maybe. Maybe not. We may discover that the steak is not as good as expected. Or we may have seen an alarming television report about mad cow disease that causes us to switch to fish. The two choice occasions were different in terms of circumstances and knowledge. Hence, they do not necessarily imply inconsistency.

An experiment may seem to reveal preference intransitivity, by showing that while X is preferred to Y , and Y is preferred to Z , Z is preferred to X . Yet this result can be explained away by showing that the three pairwise comparisons did not take place under identical conditions, or were separated in time or space. Extraneous factors may account for the apparent intransitivity. All we have to do is indicate in some way that the two Z s in the above comparisons are not quite identical. The two Z s could be slightly different in

³Hausman (1992, ch. 13) documented several attempts to explain the apparent anomalies that have been revealed by the experimenters, notably by pointing to other possible sources of utility. In some of these cases, the independence axiom is abandoned in attempts to rescue the idea of utility maximization.

timing, substance, or their informational or other contexts. We then get the result: X is preferred to Y , Y is preferred to Z_1 , and Z_2 is preferred to X . Transitivity is no longer violated.

It may be objected that if preferences are assumed stable, then evidence on revealed preference could reveal inconsistent preferences. This would however not be the case if utility depended on other factors in the environment. Consider the utility function $U = f(X, E)$, where X is a vector of consumption inputs and E is a vector of environmental or contextual conditions. Assume the function f is perfectly stable. But E can never be strictly held constant. Some part of the environment, however remotely or slightly, will inevitably alter. Hence, in practice, intransitivity (or inter-temporal inconsistency) in the rankings of the elements of vector X alone would not reveal preference inconsistency because some elements in the vector E would also have changed, even by the tiniest amount. Strictly, the environment is never constant. Consequently, because we cannot strictly and identically replicate the E conditions, intransitivity or inconsistency of X choices can never falsify the assumption of fixed preferences.

Given that we can never in principle demonstrate that some unobserved variable (like utility) is not being maximized, then the theory is invulnerable to any empirical attack. In principle, no amount of evidence can establish the non-existence of anything. Hence the standard core of expected utility theory is *unfalsifiable*.⁴

The utility-maximization assumption is unfalsifiable: but it is not a tautology (in the logical sense) because it is *conceivably false*. Tautologies and unfalsifiable propositions are often confused, but they are quite different. Logical tautologies – such as a triangle has three sides – are true by definition. By contrast, it might really be the case that an individual is not maximizing utility. Inconsistent behavior is possible in principle. But we can never establish this on the basis of empirical evidence.

This does not necessarily mean that the utility-maximization framework is useless or wrong. We do not have to uphold falsifiability as the mark of science – a criterion attributed to Karl Popper, who in fact adopted a more nuanced position (Ackerman, 1976). Neither tautological nor non-falsifiable statements are necessarily meaningless or unscientific.⁵

A key problem with utility maximization is that it is so general that it can fit anything; consequently its expressive or explanatory powers in specific

⁴Recall the Duhem-Quine thesis, which claims that it is generally impossible to falsify any single hypothesis because we always have to adopt additional hypotheses in the analysis of any set of observations (Harding, 1976). Consequently, we can never be sure that the main hypothesis is being targeted and tested on its own, and that other auxiliary hypotheses are not complicating the picture.

⁵Indeed, it is widely accepted in the philosophy of science – including by Popper – that some unfalsifiable propositions are necessary for science itself. These include the principle of determinancy (every event has a cause) and the assumption of the uniformity of nature. Without these prior assumptions, science is impossible.

instances are dramatically diminished. As Mark Blaug (1992, p. 232) put it: “[t]he rationality hypothesis by itself is rather weak. To make it yield interesting implications, we need to add auxiliary assumptions.”

Rational Choice Theory: Neither Specifically Human nor Adequate in Evolutionary Terms

The notion of utility maximization is so capacious that it goes beyond the parameters of human decision. Experimental work with rats and other animals (Kagel *et al.*, 1995) has “revealed” that animals have downward-sloping demand curves, supposedly just like humans. Gary Becker (1991, p. 307) proposed that “[e]conomic analysis is a powerful tool not only in understanding human behavior but also in understanding the behavior of other species.” Similarly, Gordon Tullock (1994) claimed that organisms – from bacteria to bears – can be treated as if they have the same general type of utility function that is attributed to humans in the microeconomics textbooks. Utility maximization can be applied to robots and the animal kingdom as well. Seemingly, we now have “evidence” of the “rationality” of everything in evolution from the amoeba onwards. As a consequence, such assumptions are expressing very little about what is specific to human nature and human society.

On its own, rational choice theory tells us nothing of significance that is specific to the human psyche, human interaction, human nature, or human society. Its very weakness, at least when applied to the human domain, stems from its excessive scope.

The non-falsifiability of the concept of rationality-as-behavioral-consistency-or-utility-maximization sustains an *epistemic* critique. It does not clinch the matter. One has also to consider the *theoretical* limitations of this stance. Here rationality-as-behavioral-consistency-or-utility-maximization falls down for at least two reasons. First, it sidelines the problem of *explaining the causes* of behavior. Gintis and Helbing are commendably explicit about this, but is a real limitation. Second, it fudges the question of the individual *development* of capacities and dispositions.

For related reasons, claims that there is an evolutionary basis for utility maximization (Robson, 2001; Gintis, 2006, p. 17) do not pass muster. It is insufficient to show that the behavioral outcomes of evolution are consistent with some utility function. Ultimately this claim is trivially true, because one can always find a function that fits.

Other defenders of rational economic man – notably Becker (1996) – treat the individual “as if” she is born with a sophisticated but fixed meta-preference function. Past economists have tolerated the “as if” neglect of real phenomena, but it no longer satisfies all scholars in this new age of exploration for evolutionary understandings of origin and development. We are interested

specifically in the human mind and human social organization. We obtain little insight in this respect from overly-capacious and unfalsifiable principles that apply to any living organism or behavioral entity.

On the basis of experimental evidence, some neuro-economists (Platt and Glimcher, 1999; Glimcher *et al.*, 2005) made the claim that the utility function exists as a physiological reality inside the brain. This claim was scrutinized by Jack Vromen (2010), who argued that at best the neurological evidence exhibits consistency with the predictions of expected utility theory. There is no evidence of actual computation of utility. Given the argument here that any observed outcomes can be made consistent with some utility function, the consistency claim is hardly powerful or surprising. But the existence claims are unsupported. Consequently, Colin F. Camerer, George Loewenstein and Drazen Prelec (2005) are skeptical of the postulate that neuro-scientific evidence supports a standard model of rational choice.

It is important to understand the nature and explain the evolutionary origins of our human dispositions. Both genetic inheritance and cultural transmission are relevant to this quest. To understand the motion of the planets or the nature of matter is to comprehend the structures and forces that lie behind events, not to imagine spirits or gods that create every eventuality. To understand human nature and society is to appreciate human dispositions and interactions, not to fit all observations of behavior to imagined mathematical functions of ever-expandable correlative capacity. Rationality in the broader sense serves an ex-post rationalization or “expression” of behavior – rather than a grounded causal explanation. A utility function may serve a limited purpose as a formalized preference ordering. Such formal constructions can have some benefits in some contexts. They can be useful shortcuts for modeling or explanatory purposes. But they do not enhance our understanding of human motivation. Utility theory is an elegant way of summarizing what we don’t know about human psychology or human evolution.

Once we join Darwin and others in the quest for an evolutionary understanding of human dispositions, defenses of preference functions as mere “expressions” of behavior are inadequate. As Gintis and Helbing sometimes do themselves, we have to go further and consider specific types of human motivation and their evolutionary origins. That is why rational choice theory is far from enough.

Morality is not Essentially a Matter of Preference

Gintis, Helbing, and their collaborators have consistently lumped “moral, social and other-regarding values” together, as if they can all be adequately “expressed” by the other-regarding preference function. However, morality, altruism and other-regarding preferences are all different things. Drawing on what I believe is the majority view among moral philosophers, I maintain

that *moral dispositions cannot be adequately summarized by any preference function.*

Of course, as noted above, preference functions can be contrived to fit any behavior. So in this shallow sense of an empirical fit, a preference function can “express” moral motivation. But it adds nothing to our understanding of our evolved moral dispositions. Furthermore, by using the language of “preference,” it degrades our understanding of the nature of morality.

Morality is complex and controversial. In Darwin’s (1871, vol. 1, pp. 87–89) account, morality results from a combination of emotional impulses and thoughtful deliberation. He argued that although primitive moral feelings have evolved for millions of years among “the progenitors of man” (1871, vol. 1, p. 162), humans alone have a developed sense of morality:

A moral being is one who is capable of comparing his past and future actions or motives, and of approving or disapproving of them. We have no reason to suppose that any of the lower animals have this capacity . . . man . . . alone can with certainty be ranked as a moral being . . .

For Darwin, morality emerged in humans upon a long-evolved foundation of instinct and impulse. Darwin also saw morality as a social phenomenon, involving social relations and shared values.

As noted above, the assumption of “other-regarding” preferences contrasts with the previously-prominent idea that economic man was entirely selfish. Yet someone with “other-regarding” preferences is still maximizing *her own* utility, and for this reason may be regarded as selfish too.

Many controversies divide moral philosophers. The best we can do here is to select a few prominent descriptions of the nature of moral judgment. The leading moral philosopher Richard M. Hare (1952) argued that morality was subject to reason and one cannot hold contradictory ethical judgments. As John L. Mackie (1977, p. 33) put it in his classic account, a moral judgment

is not purely descriptive, certainly not inert, but something that involves a call for action or for the refraining from action, and one that is absolute, not contingent upon any desire or preference or policy or choice, his own or anyone else’s.

In his very impressive philosophical account of the *Evolution of Morality*, Richard Joyce (2006, p. 70) argued on the basis of considerations in the philosophical literature that morality has most or all of the following characteristics:

1. Moral judgments express attitudes (such as approval or contempt) and also express beliefs.
2. The emotion of guilt is an important mechanism for regulating moral conduct.

3. Moral judgments transcend the interests or ends of those concerned.
4. Moral judgments imply notions of desert and justice.
5. Moral judgments are inescapable.
6. Moral judgments transcend human conventions.
7. Moral judgments govern interpersonal relations and counter self-regarding individualism.

These characteristics do not establish a *valid* morality; they instead help us to identify what is a *moral judgment*, whether acceptable or otherwise. The argument here relies on descriptive rather than normative ethics: there is no attempt here to identify the “right” morality, but instead to identify the basic nature of a moral claim. Most religions uphold moral claims, but that does not make them all right or just.

Like Darwin, Joyce emphasized the role of the emotions as well as deliberation. Joyce’s point (1) establishes that a moral judgment must involve both beliefs and sentiments, and is not reducible to either alone. If an action is impelled *purely* by emotion and sentiment then – as Darwin understood – it cannot amount to moral motivation. Deliberations and beliefs are also vital, but are themselves insufficient because they must be backed by sentiments or emotions: acting morally is more than calculated conformity to moral rules.

Moral judgments may be rationalized in various ways, but they are more than matters of propositional belief or logical syllogism. Defiance of shared moral rules in a group is often met with emotional hostility. Conformity to them may sometimes bring an emotional glow. The emotional dimension of moral rules plays an important role in their evolution and their survival. Guilt (point (2)) is a particularly important emotion that sometimes emerges after breaches of moral rules, and it too plays a part in the evolutionary process.

Joyce’s points (3) through (7) reveal the limitations of typical utility-oriented (or preference-based) approaches. Moral judgments are not simply expressions of an individual’s interests, preferences, sentiments, or beliefs; they are also claims to universality in their context, which could apply irrespective of the interests, preferences, sentiments or beliefs of those to whom they are supposed to apply.

As both Mackie and Joyce insisted, morality surpasses questions of preference. It is a matter of right or wrong, or of duty, of “doing the right thing,” irrespective of whether we like it or not. This is part of what makes us human: we are capable of considering moral rules, and understanding that their observance is more than a matter of personal whim or satisfaction. This dimension is missing in much of economics. Moral values are either ignored, or subsumed

under matters of utility or preference.⁶ This has important consequences for policy design (Hodgson, 2013).

Modern society establishes a fundamental difference between moral rules and other (normative) rules. “Murder is wrong” does not carry the same connotations as “splitting infinitives in English is wrong” or “in Britain one must drive on the left side of the road.” Linguistic and traffic rules are matters of convention; they are non-universal. But punishment may still occur when some conventions are breached. Murder is also punishable, but by contrast it is more than a breach of convention.

Threat of punishment or respect for the law are each insufficient to explain the relatively low frequency of murder and other crimes. Most of us abstain from murder not simply because the probability of severe punishment outweighs any expected benefit. Most of us refrain from murder because we believe that it is *morally wrong*; we would desist even if we lived in a country where murder went unpunished.

While there is a difference between morality and mere convention, some conventional rules seem to acquire a moral imperative when they become laws. They sometimes inherit the force of morality from other purportedly universal moral rules, particularly the need to respect others and to obey the law. While conventions may differ from culture to culture, we often conform to them, partly out of mutual respect or legal responsibility. Hence matters of mere convention can acquire some moral force if they become enshrined in law. If so, they do not necessarily become moral issues themselves, but their observance may acquire moral substance by virtue of their legal status. Consequently the moral legitimacy (or otherwise) of the legal system in the eyes of citizens is crucial (Hodgson, 2009, 2015).

It is a commonplace observation that what may be a moral rule for one culture may not be so for another. But this does not mean that moral rules are reducible to conventions. They become moral rules because many people believe in them as such, and they jointly uphold them as more than matters of convenience, self-interest or convention. The cultural specificity of some moral judgments does not justify a *normative moral relativism*, where one person’s morality is deemed as good as any other.⁷

Advocates of normative moral relativism take a further step and uphold that if an act is regarded as morally permissible in a culture then it must be deemed acceptable, even if it is regarded as wrong elsewhere. To accept such a moral relativism is to undermine an essential feature of morality itself – that it is absolute and inescapable. Because this feature is denied, such a moral relativist

⁶But some experimental economists have recently objected to subsuming every motivation or sentiment under “preferences” and argued that the behavioral evidence is better supported by richer interpretations (Wilson, 2010; Smith, 2013).

⁷The uncontroversial observation that different cultures have different ethical codes can be described as *descriptive moral relativism*. See Hodgson (2013) for a discussion.

cannot believe in *any* moral judgment in the terms defined above. Hence normative moral relativists are obliged to become moral nihilists or amoralists.

Just as there are unavoidable ontological commitments in any process of enquiry, for humans in societies there are unavoidable moral commitments. And some moral rules are ethically superior to others. Moral relativists respond that there is no way of knowing what the superior moral rules are. Yet even if there were no way of finding them, this argument is invalid. As with the epistemic error in ontology (the fact that we can never prove that the real world exists does not mean that there is not a real world outside our senses), there is a similar error in ethics (the fact that we may be unable to identify a superior moral code or prove that it is valid does not mean that a morally superior code does not exist). Ignorance of a valid morality does not mean that it is in principle unspecifiable.

Studies show a number of common features of moralities across cultures, notwithstanding other important cultural variations.⁸ All cultures regard many acts of harm against others as immoral and invest many acts of reciprocity and fairness with moral virtue. All cultures have moral rules concerning required behaviors specific to particular social positions, roles, or ranks. Moral codes restraining individual selfishness are also commonplace as well as sustaining enormous cultural diversity, genetic, and cultural co-evolution has ensured that some specific types of pro-social moral rule have endured.

In summary, moral judgments involve expressions of attitudes, beliefs and emotions but are also subject to deliberation concerning matters of fairness or justice. In contrast to standard utilitarian approaches, a moral judgment is more than mere convention; it is inescapable and transcends individual preferences or interests (Smart and Williams, 1973). A moral system refers to shared and interactively reinforced moral values in a society or social group. These definitions are incomplete and imprecise, but they are sufficient for our purposes here.

Gintis and Helbing would accept part of this. They write that the “normative aspect of social roles is motivating to social actors” of concomitant “ethical values” and “character virtues, such as honesty, trustworthiness, promise-keeping, and obedience” and “punishing others who have violated social norms even when they can gain no personal advantage thereby”. They then however claim that “[t]hese normative traits by no means contradict rationality.” Clearly, in one sense this is an empty claim, because any behavior is consistent with rational choice. But note their linking of “trait” with “rationality.” By combining the two, their previous warning that rationality is merely an expression, rather than a motivational trait, might here be violated. They claim that there is no contradiction “because individuals trade off these values against material

⁸Bok (1978), Roberts (1979), Brown (1991), Schwartz (1994), Haidt and Joseph (2004), Nichols (2004).

reward, and against each other, just as described in the economic theory of the rational actor.” Yet the assertion that “individuals trade off these values against material reward” suggests possible deliberation when faced with conflicting motivations rather than merely an “expression” of behavior. In the end, it seems in their account that rational choice theory becomes more than a mere “expression” of outcomes, and is used as part of the *explanation* of motivation and behavioral outcomes.

Morality is about “doing the right thing”: it supersedes non-moral considerations. Of course, when faced with moral dilemmas, people do often weigh up one option against the other. But to describe all this as utility maximization or a matter of preference misses the point. If people are always acting in a way that ends up maximizing *their own* utility, then they cannot be seen as truly altruistic or moral.

Indeed, it is rather odd to claim simultaneously that (a) evolution has produced individuals who always maximize their own utility and (b) they are also capable of altruism. Altruism is typically defined as costly for the individual concerned, but beneficial for others. This sits uneasily with a utility-maximizing framework. Consequently, within a utility-maximizing framework, the definition of altruism has to be constantly clarified (e.g. Bowles and Gintis, 2011). Rational choice theorists working in an evolutionary framework might awkwardly depict altruism as simultaneously involving a fitness cost and a utility gain for the agent, to preserve their near-vacuous claim that all agents are utility maximizers.

The Evolution of Morality

We require evolutionary explanations of the origin and persistence of morality. Darwin’s account of this evolution is remarkably resilient, even in the light of modern research. In his *Descent of Man*, Darwin (1871, vol. 1, p. 162) considered dispositions such as “sympathy, fidelity, and courage” that would advantage one tribe against the other in their struggle for existence, and which had been originally “acquired by the progenitors of man.” His accent on sympathy is of course reminiscent of Adam Smith in the *Theory of Moral Sentiments*. Among the other qualities, Darwin (1871, vol. 1, p. 162) listed the disposition to obey those in authority rather than follow individually-selfish motives:

Obedience . . . is of the highest value, for any form of government is better than none. Selfish and contentious people will not cohere, and without coherence nothing can be effected. A tribe possessing the above qualities in a high degree would spread and be victorious over other tribes

Obedience to authority is of course a vital mechanism in the establishment of a system of morality in society. Darwin (1871, vol. 1, p. 166) wrote:

It must not be forgotten that although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an advancement in the standard of morality and an increase in the number of well-endowed men will certainly give an immense advantage to one tribe over another. There can be no doubt that a tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to give aid to each other and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection.

Darwin proposed that groups containing individuals that devote themselves to the interests of their group will have an advantage in the struggle for survival. Among humans, binding sentiments of sympathy and solidarity are strengthened by a moral code.

Darwin's argument relied on a notion of group selection, which only recently has been rehabilitated (Sober and Wilson, 1998). This may help explain the neglect of Darwin's theory of the evolution of moral sentiments. We now have rigorous accounts of possible mechanisms group selection mechanisms, including the important distinction between genetic and cultural group selection (Henrich, 2004). The relationship between genetic and cultural factors in the evolution of morality is under dispute, but perhaps the best guess (based in current evidence) is that the genetic foundations of altruistic and moral feelings seem more likely to have evolved first through mechanisms of kin altruism and then reciprocal altruism. Then altruistic, cooperative and moral feelings required the further emergence of a culture, so that they could spread through the group and become reinforced by enduring cultural norms. In short, genetic mechanisms established critical masses of altruists in social groups, leading to the spread of cultural norms sustaining cooperation and to the development of systems of morality that further enhanced the fitness of groups (Hodgson, 2013).

When it comes to modelling the co-evolution of human dispositions and social institutions, Samuel Bowles and Herbert Gintis (2011) notably leave their rational choice framework aside, and turn instead to agent-based modelling, where instead of a single preference function, human agents are modelled in terms of modular algorithms and programmed dispositions (Bowles *et al.*, 2003). I applaud this approach. I show, furthermore, that adding properly-defined morality to these agent-based models makes a significant difference to the simulations and their outcomes (Hodgson, 2013, 2014).

Evolutionary psychologists argue that the evolution of the human brain over millions of years in social and hunter-gatherer contexts has equipped it

with specific mental modules that apply to specific problems and circumstances. Rather than an all-purpose calculator, this evolutionary perspective points to a modular brain, consisting of many special-purpose and domain-specific mechanisms (Cosmides, 1989; Cosmides and Tooby, 1994a, 1994b; Carruthers and Chamberlain, 2000; Sperber, 2005; Sarnecki, 2007).

Another way of describing these modules is *programs*. The biologist Ernst Mayr (1988, p. 45) defines “teleonomic” or “program-based” behavior as that which “*owes its goal-directedness to the operation of a program.*” Such behavior is governed by connected, rule-like dispositions. Viktor Vanberg (2002, 2004) elaborated Mayr’s argument, and showed that it provides a powerful alternative to the rationality concept.

The program-based approach relies on evolutionary theory to explain the origin of systems of rule-like dispositions, which are either inherited as instincts, or acquired as habits in a historically specific cultural setting. Generally, the human problem-solving capacity that rational choice theory attributes to “rationality” is explained in the Darwinian terms by the knowledge of the world that is incorporated in rules or programs that guide behavior. This knowledge has been accumulated through trial and error in the processes of human evolution and individual learning.

Conclusion: Rational-choice Imperialism and the End of Sociology

The paper by Gintis and Helbing is written as a theoretical framework “for sociology” but one wonders if they are doing that discipline a favor. They refer briefly to the history of sociology, noting the attempts by Vilfredo Pareto and Talcott Parsons to define its scope and to survey its frontier with economics.

Gintis and Helbing rebut Pareto with the observation that the rational choice framework can apply to the “sociological” world of “social, political, and moral values.” I take a position that is different from all three authors. The attempt by Pareto and others to define an “economic” sphere, which can be analyzed in terms of self-regarding preferences alone, is doomed. As Joseph Schumpeter (1942) and many others observed, even rapacious capitalism depends on some minimal moral values. Consequently, Pareto failed to draw a viable dividing line between the two disciplines. Gintis and Helbing have agreed on this point, but wished to subsume morality and altruism under preferences. I think that this is impossible without the evisceration of the notion of morality.

Gintis and Helbing suggest that Parsons might have thought differently if he had adequately assimilated general equilibrium theory and game theory. I am unpersuaded. Like Pareto, Parsons (1937) demarcated sociology from economics on the basis of value-motivations. But unlike Pareto he did not see the two disciplines as addressing different objects of analysis – the “economic” versus

the “social.” Parsons was deeply affected by the argument of Lionel Robbins (1932) that economics did not have a separate domain of analysis but was the “science of choice.” Parsons deftly responded, by accepting Robbins’s definition of economics as the study of choice under constraint with *given preferences*. Parsons welcomed Robbins’s claim that it was not the role of economics to explain preferences. Triumphantly raising his new flag for sociology, Parsons then claimed that it was precisely the role of sociology to explain preferences, by reference to social structures and cultural values. Parsons saw economics and sociology as investigating different aspects of the same reality, where, according to him, the domain of sociology was more expansive and less confined.

I do not think that the Parsonian research program succeeded, but that is beyond the point. If I may be allowed one unfalsifiable claim: If Parsons was brought back from the dead and he studied modern general equilibrium theory and game theory, then he would still raise the problem of explaining preferences as the distinctive task of sociology.

The most important reason why post-Parsonian sociology failed is that it could not provide an adequate explanation of the origin of preferences. Such an explanation requires reference to both the biological and cultural aspects of human evolution. But Parsons himself had been at the forefront of the successful 1930s attempts to purge sociology of all reference to biology. He was one of an influential group of social scientists that developed the mythology of “social Darwinism” in their effort to sever all connections between the social sciences and biology (Hodgson, 2004; Hodgson and Knudsen, 2010).

For sociology, the Gintis and Helbing paper is one big step forward, but another step back. By bringing in modern evolutionary theory they make a major positive contribution. But they also need to apply evolutionary theory to the origin of morality and other basic human dispositions. Instead they create a framework for sociology that effectively subsumes its concern with moral and other values under the all-embracing mantra of rationality.

The Gintis and Helbing gift to sociology is a Trojan Horse, threatening to destroy any separate core for this discipline. They bring their impressive donation within the walls, but do not acknowledge any reason for sociology’s separate existence. In this respect they are somewhat economical with the truth.

I would prefer to be more candid about the future of economics and sociology as separate disciplines. Save for the institutional inertia of journal titles and labels upon departmental doors, there is none. Since the dissolution of the tacit Robbins-Parsons agreement to divide their disciplinary territories, the marginalization of Parsons within his own discipline, and important developments within economics that redefine its scope, there is no established rationale for a division between these two disciplines. For example, “economic sociology,” claims to apply “sociological” methods to “economic” phenomena, but the nature of these methods is unexplained, and what is “economic” is unclear

(Hodgson, 2008). While I reject the total conquest of sociology by models with utility-maximizing individuals, I would limit this paradigm within economics as well. Given that, the wall between the two disciplines can come down.

I commend the rigor, elegance and evolutionary grounding of the contribution of Gintis and Helbing. But I would also warn that the attempt to subsume all human motivation or behavior within the rational choice framework threatens to undermine part of sociology's positive legacy, and it stops short of extending evolutionary theory to the explanation of moral and other motives.

References

- Ackerman, R. 1976. *The Philosophy of Karl Popper*. Amherst, MA: University of Massachusetts Press.
- Becker, G. S. 1991. *A Treatise on the Family*. 2nd edn. Cambridge, MA: Harvard University Press.
- Becker, G. S. 1996. *Accounting for Tastes*. Cambridge, MA: Harvard University Press.
- Blaug, M. 1992. *The Methodology of Economics: Or How Economists Explain*, 2nd edn. Cambridge: Cambridge University Press.
- Bok, S. 1978. *Lying: Moral Choice in Public and Private Life*. New York: Pantheon.
- Boland, L. A. 1981. "On the Futility of Criticizing the Neoclassical Maximization Hypothesis". *American Economic Review*. 71(5): 1031–1036.
- Bowles, S., J. K. Choi, and H. Astrid. 2003. "The Co-Evolution of Individual Behaviors and Social Institutions". *Journal of Theoretical Biology*. 223(2): 135–147.
- Bowles, S. and H. Gintis. 2011. *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton, NJ: Princeton University Press.
- Brown, D. E. 1991. *Human Universals*. New York: McGraw-Hill.
- Camerer, C. F., G. Loewenstein, and D. Prelec. 2005. "Neuroeconomics: How Neuroscience Can Inform Economics". *Journal of Economic Literature*. 43(1): 9–64.
- Camic, C. and G. M. Hodgson eds. 2011. *Essential Writings of Thorstein Veblen*. London and New York: Routledge.
- Carruthers, P. and A. Chamberlain eds. 2000. *Evolution and the Human Mind: Modularity, Language and Meta-Cognition*. Cambridge and New York: Cambridge University Press.
- Coase, R. H. 1977. "Economics and Contiguous Disciplines". In: *The Organization and Retrieval of Economic Knowledge*. Ed. by M. Perlman. Boulder, CO: Westview Press. 481–491.

- Cosmides, L. 1989. "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies With the Wason Selection Task". *Cognition*. 31: 187–276.
- Cosmides, L. and J. Tooby. 1994a. "Beyond Intuition and Instinct Blindness: Towards an Evolutionary Rigorous Cognitive Science". *Cognition*. 50(1–3): 41–77.
- Cosmides, L. and J. Tooby. 1994b. "Better than Rational: Evolutionary Psychology and the Invisible Hand". *American Economic Review (Papers and Proceedings)*. 84(2): 327–332.
- Darwin, C. R. 1871. *The Descent of Man, and Selection in Relation to Sex*, 2 vols. London: Murray and New York: Hill.
- Gintis, H. 2006. "Behavioral Ethics Meets Natural Justice". *Politics, Philosophy and Economics*. 5(1): 5–32.
- Gintis, H. 2007. "A Framework for the Integration of the Behavioral Sciences". *Behavioral and Brain Sciences*. 30(1): 1–16.
- Gintis, H. 2009. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton NJ: Princeton University Press.
- Glimcher, P. W., M. C. Dorris, and H. M. Bayer. 2005. "Physiologic Utility Theory and the Neuroeconomics of Choice". *Games and Economic Behavior*. 52: 213–256.
- Haidt, J. and C. Joseph. 2004. "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues". *Daedalus*. 133(4): 55–66.
- Harding, S. G., ed. 1976. *Can Theories be Refuted? Essays on the Duhem-Quine Thesis*. Dordrecht: Reidel.
- Hare, R. M. 1952. *The Language of Morals*. Oxford: Oxford University Press.
- Hausman, D. M. 1992. *The Inexact and Separate Science of Economics*. Cambridge and New York: Cambridge University Press.
- Henrich, J. 2004. "Cultural Group Selection, Coevolutionary Processes and Large-Scale Cooperation". *Journal of Economic Behavior and Organization*. 53(1): 3–35.
- Hodgson, G. M. 2001. *How Economics Forgot History: The Problem of Historical Specificity in Social Science*. London and New York: Routledge.
- Hodgson, G. M. 2004. "Social Darwinism in Anglophone Academic Journals: A Contribution to the History of the Term". *Journal of Historical Sociology*. 17(4): 428–463.
- Hodgson, G. M. 2008. "Prospects for Economic Sociology". *Philosophy of the Social Sciences*. 38(1): 133–149.
- Hodgson, G. M. 2009. "On the Institutional Foundations of Law: The Insufficiency of Custom and Private Ordering". *Journal of Economic Issues*. 43(1): 143–166.
- Hodgson, G. M. 2013. *From Pleasure Machines to Moral Communities: An Evolutionary Economics without Homo Economicus*. Chicago: University of Chicago Press.

- Hodgson, G. M. 2014. "The Evolution of Morality and the End of Economic Man". *Journal of Evolutionary Economics*. 24(1): 83–106.
- Hodgson, G. M. 2015. *Conceptualizing Capitalism: Institutions, Evolution, Future*. Chicago: University of Chicago Press.
- Hodgson, G. M. and T. Knudsen. 2010. *Darwin's Conjecture: The Search for General Principles of Social and Economic Evolution*. Chicago: University of Chicago Press.
- Joyce, R. 2006. *The Evolution of Morality*. Cambridge, MA: MIT Press.
- Kagel, J. H., R. C. Battalio, and L. Green. 1995. *Economic Choice Theory: An Experimental Analysis of Animal Behaviour*. Cambridge and New York: Cambridge University Press.
- Knight, F. H. 1933. "Preface to the Re-Issue". In: *Risk, Uncertainty and Profit*. Ed. by 2. edn. London: London School of Economics. xi–xxxvi.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.
- Marshall, A. 1885. "The Present Position of Economics". In: *Memorials of Alfred Marshall*. Ed. by A. C. Pigou. London: Macmillan. 152–174.
- Mayr, E. 1988. *Toward a New Philosophy of Biology: Observations of an Evolutionist*. Cambridge, MA and London: Harvard University Press.
- Nichols, S. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford and New York: Oxford University Press.
- Parsons, T. 1937. *The Structure of Social Action*. New York: McGraw-Hill.
- Platt, M. L. and P. W. Glimcher. 1999. "Neural Correlates of Decision Variables in Parietal Cortex". *Nature*, July 15. 400(6741): 233–238.
- Posner, R. A. 1980. "A Theory of Primitive Society, With Special Reference to Law". *Journal of Law and Economics*. 23(1): 1–53.
- Robbins, L. 1932. *An Essay on the Nature and Significance of Economic Science*. 1st edn. London: Macmillan.
- Roberts, S. 1979. *Order and Discipline: An Introduction to Legal Anthropology*. New York: St Martin's Press.
- Robson, A. J. 2001. "The Biological Basis of Human Behavior". *Journal of Economic Literature*. 39(1): 11–33.
- Samuelson, P. A. 1937. "A Note on the Measurement of Utility". *Review of Economic Studies*. 4(2): 155–161.
- Sarnecki, J. 2007. "Developmental Objections to Evolutionary Modularity". *Biology and Philosophy*. 22(4): 529–546.
- Schumpeter, J. A. 1942. *Capitalism, Socialism and Democracy, 1st edn*. London: George Allen and Unwin.
- Schwartz, S. H. 1994. "Are There Universal Aspects in the Structure and Contents of Human Values?" *Journal of Social Issues*. 50(4): 19–45.
- Sen, A. K. 1977. "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory". *Philosophy and Public Affairs*. 6(4): 317–344.

- Sen, A. K. 1987a. *On Ethics and Economics*. Oxford and New York: Basil Blackwell.
- Sen, A. K. 1987b. "Rational Behaviour". In: *The New Palgrave Dictionary of Economics*. Ed. by Eatwell, John, Milgate, Murray, Newman, and P. (eds). Vol. 4. London: Macmillan. 68–76.
- Slovic, P. and S. Lichtenstein. 1983. "Preference Reversals: A Broader Perspective". *American Economic Review*. 73(4): 596–605.
- Smart, J. J. C. and B. Williams. 1973. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.
- Smith, V. L. 2013. "Adam Smith: From Propriety and Sentiments to Property and Wealth". *Forum for Social Economics*. 42(4): 283–297.
- Sober, E. and D. S. Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
- Sperber, D. 2005. "Modularity and Relevance: How Can a Massively Modular Mind be Flexible and Context-Sensitive?" In: *The Innate mind: Structure and Contents*. Ed. by P. Carruthers, S. Laurence, and S. Stich. Oxford and New York: Oxford University Press. 53–68.
- Tullock, G. 1994. *The Economics of Non-Human Societies*. Tuscon, AZ: Pallas Press.
- Vanberg, V. J. 2002. "Rational Choice versus Program-Based Behavior: Alternative Theoretical Approaches and Their Relevance for the Study of Institutions". *Rationality and Society*. 14(1): 7–53.
- Vanberg, V. J. 2004. "The Rationality Postulate in Economics: Its Ambiguity, Its Deficiency and Its Evolutionary Alternative". *Journal of Economic Methodology*. 11(1): 1–29.
- Vromen, J. J. 2010. "On the Surprising Finding that Expected Utility is Literally Computed in the Brain". *Journal of Economic Methodology*. 17(1): 17–36.
- Weber, M. 1949. "Max Weber on the Methodology of the Social Sciences". In: *The Innate mind: Structure and Contents*. translated and edited by Edward A. Shils and Henry A. Finch. Glencoe: IL: Free Press.
- Wilson, B. J. 2010. "Social Preferences aren't Preferences". *Journal of Economic Behavior and Organization*. 73(1): 77–82.
- Winter, S. G. J. 1964. "Economic "Natural Selection" and the Theory of the Firm". 41: 225–272.