doi:10.1017/S0266267110000532

The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences, Herbert Gintis. Princeton University Press, 2009. xviii + 281 pages.

Herbert Gintis is an important contributor to a number of research programmes in the social sciences, including, but not limited to, market equilibrium theory, labour and welfare economics, experimental game theory and agent-based modelling. Gintis' work reflects an amazing breadth of knowledge of the behavioural sciences. He is ever ready to pose unusual questions and to defend unorthodox proposals. *The Bounds of Reason* is Gintis' most ambitious project to date, one that draws upon all of his extraordinary originality and learning. In this book, Gintis argues that the behavioural sciences have fallen into fundamental disarray because their practitioners employ incompatible models of human behaviour. Gintis maintains that the various behavioural sciences need to develop a unified model of choice that incorporates what we have learned from the existing choice models while eliminating their incongruities. Yet this work is much more than another clarion call for more interdisciplinary research. Gintis proposes a specific framework for creating the proposed unified model of choice, with game theory as the centrepiece. Gintis contends that game theory is the proper vehicle for analysing human decisions, and that the behavioural scientists who reject game theory do so at their own peril. Yet he also concludes that conventional game theory is not a self-sufficient theory for use in the behavioural sciences, since he maintains that apart from other elements of social theory, game theory is merely abstract mathematics with little explanatory power (pp. xiii–xiv).

Gintis identifies four different models of decision as the main models of decision currently employed in the various behavioural sciences. Gintis believes behavioural scientists tend to work with one of these models to the exclusion of the others, not appreciating that the quite different emphases of these models in fact indicate real disorder in their sciences. The biological model takes humans to be fitness-maximizing organisms whose behavioural patterns, including cooperative behavioural patterns, evolve over time (pp. 229–231). The sociological model explains human conduct in terms of societal roles regulated by social norms, where individuals serving in particular roles are motivated to comply with their role norms via a combination of material incentives and moral commitments (pp. 231–234). The psychological model explores the interrelated processes by which humans form goals, deliberate over alternative acts and learn from experience (pp. 236–238). And finally the Bayesian rational actor model of expected utility maximization gives the orthodox standard in economics for how agents should choose given coherent preferences over alternative outcomes (pp. 1–29, 234–236).

(Gintis does not use the qualifier 'Bayesian' himself in discussing the rational actor model, but I use it because it is common to refer to the maximize-expected-utility standard as the Bayesian rationality standard. Gintis (pp. 1, 234) prefers to refer to this model as the *beliefs, preferences and constraints* or *BPC model* as he believes this forestalls misunderstandings often associated with the term 'rational'.)

Gintis finds deficiencies in all four of these models. He appreciates some of the usual complaints raised against the Bayesian rationality model, including claims that experimental evidence throws this model into doubt and that this model does not reflect the bounded rationality of actual humans (pp. 235–237). But Gintis believes the most serious flaw of this model is its disregard of common beliefs across communities of individuals who interact (p. 248). Gintis attributes to most economists a background assumption he calls *methodological individualism*, the idea that social behaviour is characterized exhaustively by the characteristics and constraints of rational individuals (pp. xiv, 161). Gintis vigorously criticizes methodological individualism, and consequently believes that the Bayesian rationality model and the resulting classical game theory that are cornerstones of economics are lacking. However, Gintis also criticizes some of the other models of decision for in effect rejecting Bayesian rationality altogether. The sociological model does not assume methodological individualism. But those who use the sociological model eschew Bayesian rationality and game theory, which Gintis regards a fundamental mistake. The sociological model also lacks any mechanisms that would explain how roles and norms emerge, are transmitted, and ultimately expire in and across human communities. Research on the psychological model focuses on particularly complex decision problems of the sort that humans seldom face in life. Those who adopt the psychological model argue that Bayesian rationality might at best serve for routine choices where no ambiguities are present, but do not propose how the Bayesian model might be extended to the more complex choices they examine. Consequently, but mistakenly in Gintis' eyes, some view the Bayesian and psychological models as conceptually opposed (pp. 237–238). As for the biological model, Gintis' reservations are similar to those he has regarding the Bayesian rationality model. He argues that to the extent that practitioners of this model ignore culture, or try to reduce culture to some other notion such as reproductive fitness, they severely limit their ability to explain human behaviours. He is particularly dismayed that so many who use the Bayesian and the biological models simply ignore socialization theory. For Gintis, a particularly striking indicator of the current disorder in the behavioural sciences is the absence of any account of internalization of norms in the biological and Bayesian models (p. 234).

	Economical model	Psychological model	Sociological model	Biological model
(a) Gene–culture coveolution	R	R	R	A
(b) Sociopsychological theory of norms	R	A	A	R
(c) Game theory	A	R	R	A
(d) Bayesian rational actor model	A	R	R	A
(e) Complexity theory	R	R	R	A

A = model incorporates results from this research area.
R = model fails to incorporate results from this research area.

TABLE 1 Models of decision theory and their use of research areas according to Gintis

While he believes the conceptual divisions he perceives in the contemporary behavioural sciences are intolerable, Gintis also believes the time is ripe for a reunification. He proposes a unified model of rational decision based upon five already existing research areas: (a) gene–culture coevolution, (b) the sociopsychological theory of norms, (c) game theory, (d) the Bayesian rational actor model and (e) complexity theory (pp. 222, 247). He thinks that each of the existing four decision models fails to incorporate the results of one or more of these areas. Table 1 lists particular research areas Gintis believes the various decision models either incorporate or ignore. One can read Table 1 as a summary of an alternate version of Gintis' argument that the behavioural sciences are so muddled.

Game theory is the backbone of Gintis' analytical framework, as evidenced by the book's subtitle and his calling game theory 'The Universal Lexicon of Life' (p. 239). Indeed, Gintis devotes the bulk of *The Bounds of Reason* to discussing game theory, though he might prefer I say 'game theories', since along with the classical game theory von Neumann, Morgenstern and Nash developed in the 1940s and 1950s, Gintis refers to the more recent broad research areas connecting game theory to biology, laboratory experiments and multi-agent knowledge concepts as evolutionary game theory, behavioural game theory and epistemic game theory, respectively. Gintis' discussion of the various branches of game theory is lively, well written and illuminating. This book could even serve as an introduction to game theory for mathematically sophisticated readers. The book would serve even better readers who have some background in textbook game theory and want an overview of some

https://doi.org/10.1017/S0266267110000532 Published online by Cambridge University Press

of the most interesting recent developments in behavioural and epistemic game theory.

The Bounds of Reason presents what may be the first research proposal to connect game theory with all the major behavioural sciences. This book also summarizes a wealth of important old and new game theoretic results, some due to Gintis himself, that one will not find in any other single resource. Every social scientist interested in formal methods will find much new good food for thought here. Moral and political philosophers who take the social sciences seriously will find after reading this book they need to rethink some of their presuppositions regarding issues such as ethical egoism and the viability of cooperation in anarchy. Time will tell whether or not social scientists can successfully develop Gintis' proposed unified decision model. Social scientists may be reluctant to sign onto this project for both substantive and sociological reasons. Two of Gintis' building blocks, gene–culture coevolution and complexity theory, are so early in their development I think we cannot yet be certain they will last as distinct research areas in the long run, let alone have enduring impact on decision theory. (I say this even though I have contributed to the agent-based modelling literature that is part of complexity theory in the social sciences – see Vanderschraaf 2006, 2007, 2008.) Gintis himself thinks that the parochial character of academic disciplines will be the most serious deterrent against contributing to the unified model (p. 247). This said, Gintis has offered a compelling diagnosis of the state of theories of decision and an exciting proposal for making real progress. It is refreshing to see a substantive argument for a new and truly interdisciplinary research program. Many readers will find *The Bounds of Reason* inspiring. Some may find the work exasperating. But all who study the work seriously should find their views regarding game theory and the behavioural sciences stimulated in unexpected and fruitful ways.

The Bounds of Reason is a long and sophisticated book. Even so, Gintis does not always give a full defence of some of his claims. Gintis insists that the four existing models of decision he discusses are incompatible, but his supporting arguments are somewhat sketchy. I think that the most Gintis shows is that each of the four models is seriously deficient. One who accepts Gintis' arguments in the main should come away thinking that the four models are large and mostly nonoverlapping pieces of a much larger intellectual jigsaw puzzle, not that large parts of one or more of these models need to be discarded. Gintis stresses the importance of evolution, but includes little discussion of evolutionary game theory in this book. Gintis is well aware of this and points readers to his fine problem book, *Game Theory Evolving* (2009), which contains extensive discussion of evolutionary game theory and is a companion volume to *The Bounds of Reason* (p. xviii). Readers lacking a background in evolutionary game theory could of course also consult other works such as Maynard

Smith (1982) and Weibull (1997), but *The Bounds of Reason* is simply not a self-contained work for these readers.

Gintis also fails to address certain work where the barriers between disciplines are already dissolving. For instance, while Gintis cites certain important experiments relating neuroscience to decision making (p. 227), he includes no extended discussion of neuroeconomics and does not even mention the field by name. To be sure, neuroeconomics is new and controversial. Nevertheless, neuroeconomists are combining elements of economics and psychology in their work, and if this field is not important for Gintis' program he might have told us so and why. Another, and to my mind more serious, omission is a lack of discussion of the interplay between game theory and the social network theory that comes primarily from mathematical sociology. I would agree with Gintis that game theory is a formal lexicon of life, and that traditional game theory is seriously incomplete. Gintis believes that traditional game theorists need to incorporate mechanisms for correlating strategies, mechanisms he calls *choreographers* (pp. 41–42, pp. 132–133), as a central part of game theory, and that they have failed to do so because they have tacitly accepted methodological individualism. I agree that correlated strategies are terribly important, and will say more about them below. But traditional game theory is deficient in a second way. The traditional theory specifies the strategies players can follow together with associated payoffs, but says nothing regarding how players might choose their interaction partners. Humans tend to form ties with select members of their communities, and they tend to interact with others in proportion to how strongly they are tied to these others. Social network theory is the formal vehicle for measuring the strength of social ties. In recent years, game theorists have started to model how individuals can choose their interaction partners by embedding games into social networks. Network game theory is now a developing body of research in its own right, and one can find some of its most important results in Young (1998), Alexander (2008) and Goyal (2009). Gintis mentions social networks only in passing in *The Bounds of Reason*, but were I to revise Gintis' research programme I would add in network theory as a sixth main contributing research area.

I will close by commenting on Gintis' main proposals for reforming game theory. As noted above, Gintis identifies four game theories. He believes that in their current states, these are four largely separated theories that should be unified as part of the larger project of unifying the behavioral sciences. The branches of game theory are not so divided as Gintis suggests, in my view. Both rational choice game theorists and evolutionary game theorists have incorporated findings from behavioural game theory into their models for years. But Gintis is without doubt right to argue that epistemic game theory needs to be better integrated with

the other branches of game theory, even if he says little about how to complete this integration. Gintis is also right that doing so will reorient the entire theory. Nash equilibrium has been the central solution concept of game theory from its beginnings. But why should we suppose the players engaged in a game will follow a Nash equilibrium? Epistemic game theory gives a widely accepted answer: If the players have common knowledge of: (*i*) the game, (*ii*) their Bayesian rationality and (*iii*) their conjectures regarding each others' strategies, then their conjectures define a correlated equilibrium of the game (Aumann 1987), and if additionally (*iv*) their conjectures satisfy probabilistic independence, these conjectures define a Nash equilibrium (Aumann and Brandenburger 1995). Game theorists often proceed as if common knowledge of (*i*), (*ii*) and (*iv*) are unproblematic, so that the only real challenge is accounting for common knowledge of conjectures in order to predict *which* Nash equilibrium the players will follow. A few previous authors have objected that there is no a priori reason to suppose that players' conjectures satisfy probabilistic independence. Strikingly, Gintis argues that game theorists have too casually accepted common knowledge of rationality and the game, which is the basis of the rationalizability solution concept. He maintains that the experimental evidence should lead us to doubt that there are good a priori reasons to assume even this much common knowledge. Gintis insists that epistemically, common knowledge of the game and of rationality is actually an *event* the players infer from the outcomes they follow, same as they might infer common knowledge of conjectures under the right circumstances (pp. 100–101, p. 117). I find this the most original and interesting specific conclusion in this book.

Gintis follows other authors who characterize the norms that people follow in social life as equilibria of appropriate games (Sugden 2004; Binmore 2005; Bicchieri 2006), although Gintis adds an additional twist by arguing that individuals need a *normative disposition* to follow these equilibria (p. 133). For Gintis, a norm serves as a choreographer that specifies which strategies players are to follow, and rational players follow their prescribed strategies because the norm also supplies the epistemic conditions for common prior probabilities over the relevant states of the world, including the possible outcomes of the game (p. 133). This idea of course builds upon and generalizes David Lewis' analysis of a convention as a coordination equilibrium of a game that players follow because they have common knowledge that they follow this equilibrium, rather than any other (1969). Given his analysis, Gintis argues that methodological individualism is descriptively false, since the choreographer that guides conduct in actual human communities is a correlating mechanism external to the players and their strategies and payoffs. Gintis also argues on account of his analysis that correlated equilibrium should be the central

solution concept in game theory, since common knowledge of priors implies the players' conjectures are in correlated equilibrium. I agree with Gintis that norms are best characterized as correlated equilibria. From stopping at red lights to conveying property rights of unowned goods to first finders (pp. 135–136), the norms we follow are rules for following strategy systems tied to clues in our environment, and these systems are formally correlated equilibria. I also agree with Gintis that game theorists should pay far greater attention to the correlated equilibrium concept, although I base my opinion more upon my doubts regarding the probabilistic independence assumption of the Nash equilibrium concept than upon Gintis' analysis of norms.

Still, I think Gintis' analysis of norms raises serious questions, which point to possibilities for future research. First, just how do the members of a community acquire the common knowledge associated with a given norm? Gintis gives a fine review of the usual explanation of common knowledge in terms of public events, but a typical norm regulates a large community with changing membership over a long stretch of time. Such a norm simply cannot be promulgated among everyone who is to follow it in a public announcement, so community members must acquire the common knowledge that underwrites this norm some other way, if at all. One possible approach to accounting for common knowledge of norms is motivated by the pioneering studies of Boyd and Richerson (1985, 2005), who analyse the transmission of culture through the lens of evolutionary theory. This approach, which I think is the most promising, has yet to be integrated into epistemic game theory.

Here is a second question for Gintis: Why do the members of a community follow a particular choreographer, and no other? Why assign property rights over a previously unowned good to the claimant who found the good first, and not the oldest claimant, or the tallest claimant, or . . . ? A natural answer to this question of equilibrium selection, and one I think Gintis would favour, again appeals to evolution: a particular rule becomes the choreographer for a community via some dynamical adjustment process. But this sort of answer raises yet another serious question. An indefinitely repeated game has infinitely many different pure strategies, so any evolutionary analysis of the game must first set limits on the number of strategies that could evolve. How are the rules that are viable candidates for the choreographer to be singled out for evolutionary analysis? I think the best answer to this new question is that certain strategies will 'stand out' for the members of a community, so that they will be willing to try out these strategies in a process of trial and error learning that can result in some of them evolving into the strategies of community norms. This is analogous to the idea that certain equilibria in a coordination game are somehow *focal* or *salient* (Schelling 1960; Lewis 1969; Sugden 2004). But a salience explanation evidently

presupposes reciprocal expectations in the community regarding which strategies 'stand out', so we need some common knowledge or something like it in order to carry out an evolutionary analysis of the evolution of norms. In the last paragraph I noted that one might use evolution as part of an explanation of reciprocal expectations, while here I have just allowed that reciprocal expectations might support evolutionary analysis. But perhaps this circle is not vicious. As I suggested in the preceding paragraph, game theorists have yet to fully integrate the evolutionary and the epistemic approaches. I think this is not surprising, since epistemic game theory assumes that players have certain similar epistemic powers and explores what such players can infer about each other, while the existing evolutionary game theory models analyse how strategies can spread in populations with little if any reference to what individuals in these populations might know. Again, Gintis does not say a lot in *The Bounds of Reason* regarding how social scientists might bridge the gap between the epistemic and evolutionary game theories, but bridging this will be an especially important and challenging part of completing his programme.

Peter Vanderschraaf
University of California Merced

REFERENCES

Alexander, J. M. 2008. *The Structural Evolution of Morality*. Cambridge: Cambridge University Press.
Aumann, R. 1987. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica* 55: 1–18.
Aumann, R. and A. Brandenburger 1995. Epistemic conditions for Nash Equilibrium. *Econometrica* 63: 1161–1180.
Bicchieri, C. 2006. *The Grammar of Society*. Cambridge: Cambridge University Press.
Binmore, K. 2005. *Natural Justice*. Oxford: Oxford University Press.
Boyd, R. and P. Richerson 1985. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.
Boyd, R. and P. Richerson 2005. *The Origin and Evolution of Cultures*. New York: Oxford University Press.
Gintis, H. 2009. *Game Theory Evolving*. Princeton: Princeton University Press.
Goyal, S. 2009. *Connections: An Introduction to the Economics of Networks*. Princeton: Princeton University Press.
Lewis, D. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press. Maynard
Smith, J. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
Schelling, T. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
Sugden, R. 2004. *The Economics of Rights, Cooperation and Welfare*, 2nd Edn. New York: Palgrave MacMillan.
Vanderschraaf, P. 2006. War or Peace?: A dynamical analysis of anarchy. *Economics and Philosophy* 22: 243–279.

https://doi.org/10.1017/S0266267110000532 Published online by Cambridge University Press

Vanderschraaf, P. 2007. Covenants and reputations. *Synthese* 157: 167–195.
Vanderschraaf, P. 2008. Game theory meets threshold analysis: reappraising the paradoxes of anarchy and revolution. *British Journal for Philosophy of Science* 59: 1–39.
Weibull, J. 1997. *Evolutionary Game Theory*. Cambridge, MA: MIT Press.
Young, H. P. 1998. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton: Princeton University Press.