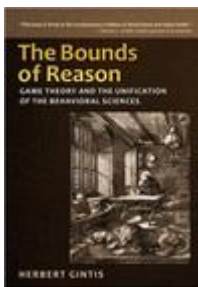


Games, Norms, and the Behavioral Sciences

A review of



The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences

by Herbert Gintis

Princeton, NJ: Princeton University Press, 2009. 286 pp. ISBN

978-0-691-14052-0. \$35.00



Reviewed by

[Gordon Pitz](#)

Game theory is a mathematical analysis of strategic interactions. Its broad principles have been understood for 300 years or more, but not until the theory was formalized by von Neumann and Morgenstern (1944) did it become an important tool for social scientists. It acquired its modern form when John Nash (1950) proved his theorem concerning the existence of the Nash equilibrium.

At first it appeared that game theory would provide a complete account of the behavior of a rational person and that most intelligent people would behave that way. It soon became clear, though, that there are serious limitations to its predictive power and that it is inadequate as a prescription for rationality (see Colman, 2003).

Consider the Prisoner's Dilemma game. As is well known, the Nash equilibrium strategy in this game is for both players to defect, although by cooperating they could both do better. The Ultimatum game creates a rather different paradox. One player is allotted an amount of money; he or she decides how to divide it with a second player. If the second

player accepts the arrangement, they each keep their portion. If the second player rejects the arrangement, neither receives anything. Game theory prescribes that the first player should offer the smallest amount possible (say, \$1.00), and the second player should accept it (after all, \$1.00 is better than nothing). Although chimpanzees apparently behave this way (Jensen, Call, & Tomasello, 2007), humans generally do not.

In the High–Low or Merchants’ Wares game, two players each choose either H or L. If both choose H, they receive, say, \$10.00. If both choose L, they receive, say, \$5.00. If their choices are different, neither receives anything. The players maximize their winnings if both choose H, and most pairs of players do this. Yet the surprising property of this game is that nothing in game theory requires or justifies either player’s choosing H.

Using Social Theory to Supplement Game Theory

The examples illustrate the importance of factors such as cooperation, fairness, and trust for explaining strategic behavior. In *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*, Herbert Gintis proposes an extension of game theory integrating the classical concept of rationality with these social factors. This extension, he argues, can effectively solve difficulties of the kind that have been documented.

Gintis begins by emphasizing the importance of game theory in any account of human behavior: “Social theory without game theory is a handicapped enterprise” (p. xiii). For game theory to serve any practical value, however, it must be integrated with other theoretical concepts: “Game theory without broader social theory is merely technical bravado” (p. xiii). His point is that the Nash equilibrium is not an appropriate construct for social theory. Traditional applications of game theory rely on such assumptions as “common knowledge of rationality” (CKR)—each player is assumed to be rational, to know that the other is rational, to know that the others know the others are rational, and so on. Gintis finds such assumptions unsatisfactory. They may or may not hold in a given context and thus cannot be a foundation for prescribing or predicting behavior.

The solution proposed by Gintis is a device that coordinates the behavior of the players in a game, a device he calls the *choreographer*. The choreographer is not a person but a set of social norms, emergent properties of social systems. These norms may be learned by individuals who are genetically predisposed to recognize and obey them, and people will then tend to obey the norms even when it is costly to do so.

Consider how this might work in the High–Low game. As noted, most players have no difficulty in achieving the maximum payoff. This must require common conjectures by each player concerning the other’s strategy. Classical game theory cannot explain where these common conjectures come from, but when we introduce the idea of social norms, then

the solution emerges. For Gintis, the source of these norms is the gene–culture evolutionary history that players share.

The choreography provided by this common evolutionary history does not alone determine behavior. Individuals can be both self-regarding and other-regarding. Game theory has typically emphasized self-regarding tendencies. Gintis seeks to supplement the theory with character virtues (honesty, trust, etc.), altruistic cooperation (helping others at a cost to oneself), and altruistic punishment (punishing others at a cost to oneself). These other-regarding concerns are partial determinants of a person's preferences, although they can be traded off against self-regarding concerns: "If the cost of virtue is sufficiently high, and the probability of detection of a breach of virtue is sufficiently small, many individuals will behave dishonestly" (p. 73).

Gintis's theory is founded on the concept of a rational actor, but his definition of rationality is a modest one and requires only consistency of preferences. He argues that preference consistency is a necessary consequence of evolutionary biology. Nevertheless, a number of psychologists have argued that consistency is often violated and that preferences are influenced in a significant way by framing effects (see, e.g., Dawes, 2001). Gintis is not dismayed by these demonstrations. Some he finds trivial. Others, he argues, can be resolved by introducing contextual factors into the decision maker's preference function.

While individual rationality is a necessary assumption, it is not sufficient. Gintis reviews a number of topics in game theory that illustrate the theory's failure to explain human behavior. He argues that the source of these failures is *methodological individualism*, the belief that we can analyze strategic behavior simply by making assumptions about the individual rationality of the agents. The CKR assumption described earlier is an example of this individualism.

The alternative approach is to assume that societies have emergent properties that cannot be reduced to features of the individuals. These emergent properties are social norms that have emerged from our evolutionary history and provide the coordination that optimizes the outcome of human interactions.

To show how this might work, Gintis devotes a chapter to the analysis of property rights, illustrating the interaction of evolutionary pressures, social norms, and strategic interaction. A well-established finding in behavioral decision theory is loss aversion and the endowment effect. People place a higher value on a good when they possess it than when they do not possess it; thus they are more averse to its loss than might be predicted by their willingness to purchase it.

Gintis's explanation for the finding uses an evolutionary analysis of property rights, showing that under certain conditions a person, or any territorial animal, would be willing to commit more resources in defense of his or her property or territory than would an intruder who wants to take the property. Thus, the endowment effect emerges as an example of a choreographing device that enables members of a social species to solve otherwise intractable problems.

Game Theory and the Unification of the Behavioral Sciences

Gintis has a more ambitious goal in mind than extending game theory. He sees his theory as a central pillar in a unified social and behavioral science. In strong terms, he decries the separate development of the various behavioral sciences: “How can the basic conceptual frameworks . . . , as outlined by their respective Great Masters and as taught to Ph. D. candidates, have almost nothing in common?” (p. xv). He is concerned that “the behavioral sciences currently consist of some fields where theory has evolved virtually without regard for the facts and others where facts abound and theory is absent” (p. xvi), and he appears to seek nothing less than the overthrow of established disciplinary boundaries.

His argument for unification is presented with a hint of paranoia:

I propose . . . a conceptual integration . . . that is analytically and empirically defensible and could be implemented now were it not for the virtually impassible feudal organization of the behavior disciplines . . . , the structure of research funding agencies . . . and interdisciplinary ethics that value comfort and tradition over the struggle for truth. (pp. xv–xvi)

This reads like the manifesto of a revolutionary, not the cautious argument of an academic scholar.

The conceptual integration, introduced with such strong language, is based on the five units that Gintis has woven together in the remainder of the book: gene–culture co-evolution, the theory of norms, game theory, the rational actor model, and complexity theory. He illustrates the ability of the integration to account for behavior in a variety of contexts and insists that it does not require a revolutionary change in any discipline. But does this really justify destruction of the organizations that administer these disciplines?

The strongest objections to Gintis’s proposal will be prompted by his dismissal of some approaches that are popular in individual disciplines. He is especially harsh, for example, in his treatment of psychologists who have criticized the rational actor model. “This dismissal of traditional decision theory may be emotionally satisfying, but it is immature, short-sighted, and scientifically destructive” (p. 246). Comments such as this are unlikely to encourage interdisciplinary cooperation. While several of the criticisms of rationality that have been raised by psychologists can be handled by his theory, others, such as preference reversal findings, he simply dismisses as inconsequential. The off-hand dismissal does not strengthen his overall case.

Conclusion

Gintis has set an ambitious goal for himself. Whether he has achieved that goal will not be known until others have a chance to examine his proposals carefully and find out how well they can account for the complexities of human behavior. Researchers in the social and behavioral sciences will find his book useful and thought provoking. A program of research based on Gintis's principles might be very productive. The book should be read by anyone who seeks to understand choice behavior and strategic interactions, even if he or she disagrees with his conclusions.

Gintis's presentation is heavily mathematical. The mathematics may provide the logical foundation for some of his theoretical assertions, but many readers will be intimidated by the pervasive hieroglyphics. If necessary, readers should skip past the mathematics, for it is not essential in understanding his main ideas.

Although each chapter provides a carefully developed treatment of a specific topic, the transitions within and across chapters are sometimes hard to follow. In fact, I strongly recommend that readers start with the final chapter, where in a little more than a page Gintis provides a summary of his main points. The remainder of the book is a lot easier to follow once one understands the overall plan.

Gintis's argument that game theory is an indispensable tool for modeling human behavior is convincing, and his integration of broader themes from social theory is a valuable extension of game theory. Even though I found his rejection of criticisms of the rational actor model to be rather glib, his comprehensive theory of behavior is worth further assessment. Game theory appears to be a powerful method for formalizing altruism and the character virtues, concepts that might otherwise appear to be imprecise and nebulous.

His framework for a unified approach to the behavioral sciences may well lead to worthwhile innovations. In spite of his strongly worded manifesto, though, I think it unlikely that the book will lead to the unification he is seeking. Gintis may be correct in blaming institutional obstacles on "the semifeudal nature of modern behavioral disciplines" (p. 247), but I suspect that he is seeking to overturn the same social forces he describes in his own theory. He may have to be satisfied with a more gradual evolution of the norms that govern behavior in the behavioral sciences.

References

- Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *The Behavioral and Brain Sciences*, 26, 139–153.

- Dawes, R. (2001). *Everyday irrationality*. Boulder, CO: Westview Press. PsycINFO
- Jensen, K., Call, J., & Tomasello, M. (2007, October 5). Chimpanzees are rational maximizers in an ultimatum game. *Science*, *318*, 107–109. PsycINFO Article
- Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America*, *36*, 48–49.
- von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
-