Preface

The eternal mystery of the world is its comprehensibility.

Albert Einstein

Mathematics without natural history is sterile, but natural history without mathematics is muddled.

John Maynard Smith

Game theory is central to understanding the dynamics of life forms in general, and humans in particular. Living creatures not only play games, but dynamically transform the games they play, and have themselves thereby evolved their unique identities. For this reason, the material in this book is foundational to all the behavioral sciences, from biology, psychology and economics to anthropology, sociology, and political science. Disciplines that slight game theory are the worse—indeed, much worse—for it.

We humans have a completely stunning capacity to reason, and to apply the fruits of reason to the transformation of our social existence. Social interactions in a vast array of species can be analyzed with game theory, yet only humans are capable of playing a game after being told its rules. This book is based on the appreciation that evolution and reason interact in constituting the social life and strategic interaction of humans.

Game theory, however, is not everything. This book systematically refutes one of the guiding prejudices of contemporary game theory. This is the notion that game theory is, insofar as human beings are rational, sufficient to explain all of human social existence. In fact, game theory is complementary to ideas developed and championed in all the behavioral disciplines. Behavioral scientists who have rejected game theory in reaction to the extravagant claims of some of its adherents may thus want to reconsider their position, recognizing the fact that, just as game theory without broader social theory is merely technical bravado, so social theory without game theory is a handicapped enterprise.

The reigning culture in game theory asserts the sufficiency of game theory, allowing game theorists to do social theory without regard either for the facts or the theoretical contributions of the other social sciences. Only

the feudal structure of the behavioral disciplines could possibly permit the persistence of such a manifestly absurd notion in a group of intelligent and open-minded scientists. Game theorists act like the proverbial "man with a hammer," for whom "all problems look like nails." I have explicitly started this volume with a broad array of social facts drawn from behavioral decision theory and behavioral game theory to disabuse the reader from this crippling notion. Game theory is a wonderful hammer, indeed a magical hammer. But, it is only a hammer, and not the only occupant of the social scientist's toolbox.

The most fundamental failure of game theory is its lack of a theory of when and how rational agents share mental constructs. The assumption that humans are rational is an excellent first approximation. But, the Bayesian rational actors favored by contemporary game theory live in a universe of subjectivity, and instead of constructing a truly social epistemology, game theorists have developed a variety of subterfuges that make it appear that rational agents may enjoy a commonality of belief (common priors, common knowledge), but all are failures. Humans have a *social epistemology*, meaning that we have reasoning processes that afford us forms of knowledge and understanding, especially the understanding and the sharing of the content of other minds, that are unavailable to merely "rational" creatures. This social epistemology characterizes our species. The bounds of reason are thus not the irrational, but the social.

That game theory does not stand alone entails denying *methodological individualism*, a philosophical position asserting that all social phenomena can be explained purely in terms of the characteristics of rational agents, the actions available to them, and the constraints that they face. This position is incorrect because, as we shall see, human society is a system with *emergent properties*, including social norms, that cannot be analytically derived from a model of interacting rational agents, any more than the chemical and biological properties of matter can be analytically derived from our knowledge of the properties of fundamental particles.

Evolutionary game theory often succeeds where classical game theory fails (Gintis 2009). The evolutionary approach to strategic interaction helps us understand the emergence, transformation, and stabilization of behaviors. In evolutionary game theory, successful strategies diffuse across populations of players rather than being learned inductively by disembodied rational agents. Moreover, reasoning is costly, so rational agents will not often even attempt to learn optimal strategies to complicated games, but

rather will copy the behavior of successful agents whom they encounter. Evolutionary game theory allows us to investigate the interaction of learning, mutation, and imitation in the spread of strategies when information processing is costly.

But, evolutionary game theory cannot deal with unique events, such as strangers interacting in a novel environment, or Middle East peace negotiations. Moreover, by assuming that agents have very low level cognitive capacities, evolutionary game theory ignores one of the most important of human capacities, that of being able to reason. Human society is an evolved system, but human reason is one of the key evolutionary forced involved. This book champions a unified approach based on modal logic, epistemic game theory, and social epistemology as an alternative to classical and a supplement to evolutionary game theory.

This approach holds that human behavior is most fruitfully modeled as the interaction of rational agents with a social epistemology, in the context of social norms that act as correlating devices that *choreograph* social interaction. This approach challenges contemporary sociology, which rejects the rational actor model. My response to the sociologists is that this rejection is the reason sociological theory has atrophied since the death of Talcott Parsons in 1979. This approach also challenges contemporary social psychology, which not only rejects the rational actor model, but generally delights in uncovering human "irrationalities." My response to the social psychologists is that this rejection accounts for the absence of a firm analytical base for the discipline, which must content itself with a host of nano-models that illuminate highly specific aspects of human functioning with no analytical linkages among them.

The self-conceptions and dividing lines among the behavioral disciplines make no scientific sense. How can there be three separate fields, sociology, anthropology, and social psychology, for instance, studying social behavior and organization? How can the basic conceptual frameworks for the three fields, as outlined by their respective Great Masters and as taught to Ph.D. candidates, have almost nothing in common? In the name of science, these arbitrarities must be abolished. I propose, in the final chapter, a conceptual integration of the behavioral sciences that is analytically and empirically defensible, and could be implemented now, were it not for the virtually impassible feudal organization of the behavior disciplines in the contemporary university system, the structure of research funding agencies

that mirror this feudal organization, and inter-disciplinary ethics that value comfort and tradition over the struggle for truth.

Game theory is a tool for investigating the world. By allowing us to specify carefully the conditions of social interaction (player characteristics, rules, informational assumptions, payoffs), its predictions can be *tested*, and the results can be replicated in different laboratory settings. For this reason, *behavioral game theory* has become increasingly influential in affecting research priorities. This aspect of game theory cannot be overstressed, because 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.

Economic theory has been particularly compromised by its neglect of the facts concerning human behavior. This situation became clear to me in the summer of 2001, when I happened to be reading a popular introductory graduate text in quantum mechanics, as well as a leading graduate text in microeconomics. The physics text began with the anomaly of black body radiation, which was inexplicable using the standard tools of electromagnetic theory. In 1900, Max Planck derived a formula that fit the data perfectly, assuming that radiation was discrete rather than continuous. In 1905, Albert Einstein explained another anomaly of classical electromagnetic theory, the photoelectric effect, using Planck's trick. The text continued, page after page, with new anomalies (Compton scattering, the spectral lines of elements of low atomic number, etc.) and new, partially successful models explaining the anomalies. This culminated in about 1925 with Heisenberg's wave mechanics and Schrödinger's equation, which fully unified the field.

By contrast, the microeconomics text, despite its beauty, did not contain a single fact in the whole thousand page volume. Rather, the authors build economic theory in axiomatic fashion, making assumptions on the basis of their intuitive plausibility, their incorporation of the "stylized facts" of everyday life, or their appeal to the principles of rational thought. A bounty of excellent economic theory was developed in the Twentieth century in this manner. But, the well has run dry. We will see that empirical evidence challenges the very foundations of both classical game theory and neoclassical economics. Future advances in economics will require that model-building dialogue with empirical testing, behavioral data-gathering, and agent-based models.

A simple generalization can be made: decision theory has developed valid algorithms by which people can best attain their objectives. Given these

objectives, when people have the informational prerequisites of decision theory, yet fail to act as predicted, the theory is generally correct and the observed behavior faulty. Indeed, when deviations from theoretical predictions are pointed out to intelligent individuals, they generally agree that they have erred. By contrast, the extension of decision theory to the *strategic interaction* of Bayesian decision-makers has led to a limited array of useful principles, and when behavior differs from prediction, people generally stand by their behavior.

Most users of game theory remain unaware of this fact. Rather, the contemporary culture of game theory (as measured by what is accepted without complaint in a journal article) is to act as if epistemic game theory, which has flourished in the past two decades, did not exist. Thus, it is virtually universal to assume that rational agents will play mixed strategies, use backward induction, and more generally, play a Nash equilibrium. When people do not conform to these expectations, their rationality is called into question, whereas in fact, none of these assumptions can be successfully defended. Rational agents just do not behave the way classical game theory predicts, except in certain settings, such as anonymous market interactions.

The reason for the inability of decision theory to extend to strategic interaction is quite simple. Decision theory shows that when a few plausible axioms hold, we can model agents has having beliefs (subjective priors), and a utility function over outcomes such that the agent's choices maximize the expected utility of the outcomes. In strategic interaction, there is nothing guaranteeing that all interacting parties have mutually consistent beliefs. Yet, as we shall see, a high degree of inter-subjective belief consistency is required to ensure that agents will play coordinated (Nash and correlated equilibrium) strategies.

The behavioral sciences have yet to adopt a serious commitment to linking basic theory and empirical research. Indeed, the various behavioral disciplines hold distinct and incompatible models of human behavior, yet their leading theoreticians make no attempt to adjudicate these differences (see chapter 12). Within economics there have been stunning advances in both theory and empirical data in the past few decades, yet theoreticians and experimentalists retain a hostile attitude to each other's work. This bizarre state of affairs must end.

It is often said that the mathematical rigor of contemporary economic theory is due to the economists' "physics-envy." In fact, physicists generally judge models according to their ability to account for the facts, not their

mathematical rigor. Physicists generally believe that rigor is the enemy of creative physical insight, and they leave rigorous formulations to the mathematicians. The economic theorists' overvaluing of rigor is a symptom of their undervaluing of explanatory power. The truth is its own justification, and needs no help from rigor.

Game theory can be used very profitably by researchers who do not know or care about mathematical intricacies, but rather treat mathematics as but one of several tools deployed in the search for truth. I assert, then that my arguments are correct and logically argued. I will leave rigor to the mathematicians.

In a companion volume, *Game Theory Evolving* (2009), I stress that understanding game theory requires solving lots of problems. I also stress therein that many of the weaknesses of classical game theory have beautiful remedies in evolutionary game theory. Neither of these considerations is dealt with in *The Bounds of Reason*, so I invite the reader to treat *Game Theory Evolving* as a complementary treatise.