

Experimental Economics Will Foster a Renaissance of Economic Theory

Herbert Gintis*

August 28, 2008

What does economic theory assert? To what extent are its assertions confirmed by experiment? Where theoretical assertions are not confirmed, are there adjustments that expand the explanatory power of the theory and bring it into harmony with experiment?

Vernon Smith, in his penetrating essay, “Theory and Experiment,” draws on his broad knowledge of experiments to suggest that subject behavior often diverges from game theoretic predictions, and in many cases there are no known correctives. The subtitle of his essay, “What are the Questions?” suggests that minor fudging with theory is unlikely to solve our problems because “circumstances unknown to us” are responsible for the failure of contemporary economic theory.

This is precisely the set of issues with which I deal in *The Bounds of Reason* (Princeton University Press, 2009). My remarks here draw upon several themes from this book relevant to specifying a modified research agenda for the study of human strategic interaction.

The core of economic theory is the rational actor model, which holds that individuals have preferences over outcomes, they have beliefs as to how choices affect the relative probability of various outcomes (called “subjective priors”), and they maximize their expected payoffs given these preferences and beliefs, subject to whatever material, informational, and other constraints they face. Although many observers consider the experimental work of Daniel Kahneman and others (Kahneman, Slovic and Tversky 1982, Gigerenzer and Todd 1999) as destructive of the rational actor model, in fact, their research has considerably strengthened its explanatory power, albeit at the expense of increasing its complexity (Gintis 2009a, Chs. 1,12). Most important, these studies have shown the importance of

*Santa Fe Institute and Central European University. The following material is adapted from my forthcoming book, *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences* (Princeton University Press, 2009). I would like to thank the European Science Foundation for research support.

using heuristics to conserve on information processing, and of including the current state (physiological, temporal, ownership) of the individual as an argument of the preference ordering.

In the 1980's, the core of economic theory was extended to include game theory, viewed as the theory of the strategic interaction of Bayesian rational actors. Game theory has become the centerpiece of microeconomic theory, displacing traditional Marshallian and Walrasian models of production, consumption, and exchange for the working economist. It is important to note, however, that this exalted position is quite recent. The textbooks that fostered the received wisdom of today's economists were written in the early years of the game theory revival, between 1988 and 1995. These textbooks include misconceptions promulgated in the 1980's, that have been transmitted to the current generation of economists. As I shall show, some of these misconceptions have been repaired in more recent contributions to game theory. Contemporary experimentalists, however, generally rely on the more traditional body of outdated theory.

In this paper, I rely on more recent theoretical contributions, especially those of Robert Aumann and Adam Brandenburger and their coworkers, to bring game theory up to date. I then argue that, however powerful, game theory and the rational actor model alone are simply not enough to explain human strategic interaction. Moreover, because individuals bring the complexities of society into even the controlled conditions of the laboratory, game theory and the rational actor model are insufficient to explain experimental results. Concretely, I suggest that (a) these analytical tools must be supplemented by what I term the *psycho-social theory of norms*; (b) the *correlated equilibrium* should replace the Nash equilibrium as the central equilibrium concept of game theory; (c) *social norms* should be recognized as objective correlating devices in instantiating correlated equilibria; and (d) the human predisposition to *internalize social norms*, and more generally to include ethical values as arguments in personal preference orderings, should be included in analytical models of strategic interaction.

1 From Game Theory to Social Theory

Early work on game theory culminated in Luce and Raiffa's (1957) tour-de-force, after which interest in game theory abated. Renewed interest was sparked in the 1980's, the year 1982 alone seeing the publication of Rubinstein's famous game-theoretic model of bargaining, Milgrom and Weber's theory of auctions, Bengt Holmström's game-theoretic microfoundations of organizations, and the famous "gang of four" (Kreps, Milgrom, Roberts, and Wilson) explanation of cooperation in the finitely repeated prisoner's dilemma. Game theory became the fundamental

microeconomic approach in the following decade, with textbooks by Tirole (1988), Rasmusen (1989), Kreps (1990), Myerson (1991), Fudenberg and Tirole (1991), Osborne and Rubinstein (1994) and culminating in the current industry standard Mas-Colell, Whinston and Green (1995).

These textbooks promote two basic principles that are now accepted uncritically by most economists, and have become the profession's received wisdom. The first principle is that rational agents play subgame perfect Nash equilibria. This principle, which I shall argue is simply not true. As I show in Section 2, game theory in fact does *not* imply that rational agents play a subgame perfect Nash equilibrium, even when such an equilibrium is the unique equilibrium of the game. The conditions under which rational agents choose subgame perfect Nash equilibria is in fact an important, but unsolved problem.

This message has not filtered through to the textbooks, and hence is ignored by most economists, who rarely read the primary sources. Not surprisingly, when experiments showed that subjects often do not play subgame perfect Nash equilibria (Camerer 2003), the natural response is that subjects are not rational. These experimental findings are reviewed by Vernon (Smith 2009), who wisely rejects this facile answer, and bids that we search for something more substantive.

The second basic principle promoted by the textbooks is that all social phenomena can be modeled as Nash equilibria of an appropriately specified game played by rational agents. This principle is never stated explicitly, but rather is a form of *tacit knowledge* (Polanyi 1966) the student infers from the fact that no social construct other than the formal rules of the game, and no properties of the players other than their preferences and subjective priors, are admitted as basic causal factors or unexplained *explananda*. Yet, this form of *methodological individualism* is never explicitly defended.

In Section 3, I suggest that methodological individualism is incorrect. At least one additional social construct above the level of the individual must be added to explain individual behavior, that of the *social norm*, and at least one human psychological characteristic must be added, that of an understanding of and predisposition towards conforming to social norms. The social theory linking social norms and individual behavior is well developed, and may be called the *psycho-social theory of norms*.

According to this theory, termed *role theory* in sociology (Linton 1936, Parsons 1967), upon encountering a social interaction, individuals first infer from social cues the nature of the interaction and deduce the social norms appropriate to this interaction (this is why Smith's Assumption 6 is incorrect—context matters in addition to the underlying abstract game structure). Individuals then use this information to constitute their *beliefs* concerning the likely behaviors of others on the one hand, the payoffs they attach to alternative actions, and the behavior appro-

priate to role-performance. Moreover, they use this information to constitute their *preferences* over these payoffs, because human agents have a socially constituted genetic predisposition to treat conformity to legitimate social norms as personally valuable, and hence represented in their preference orderings.

The concept of social norms is not absent from standard game theory. Several researchers have developed the notion that social norms are Nash equilibria of social games (Lewis 1969, Binmore 2005, Bicchieri 2006). This approach, despite the many insights it offers, nevertheless remains bound to methodological individualism: social norms are explained as the product of the strategic interaction of rational agents. The psycho-social theory of norms goes beyond this to claim that social norms are not simply *coordinating* devices, but also *motivating* devices, inducing individuals to sacrifice on behalf of compliance with norms because they are intrinsically valued. In this manner, social life is imbued with an ethical dimension absent from standard game theory.

It follows from the psycho-social theory of norms that individuals' probability distributions over states of nature and their preferences are not the purely personal characteristics (subjective priors and preference orderings) of the standard rational actor model, but rather are the product of the interaction of personal characteristics and the social context. This is why experimentalists have had such difficulty in modeling strategic interaction: the parameters of the preference function are situation-dependent.

According to the standard economics model, rational actors should be self-regarding unless expressing social preferences allows them to build reputations for cooperation in the future. An overwhelming body of experimental evidence supports the fact that individuals exhibit non-self-regarding preferences even in one-shot anonymous interactions. The standard interpretation in behavioral game theory of this ostensibly bizarre behavior is that subjects have a cognitive deficit. Even Vernon Smith, who generally prefers to avoid this interpretation, says "The bottom line appears to be that the abstract concept of single play invokes conditions sufficiently remote from much human experience that it may be operationally difficult to penetrate." In fact, unselfish behavior in one-shot interactions are commonplace in daily life; life in modern society would be intolerable but for the kindness of strangers, and most of us go to great lengths in public to avoid incurring even a disapproving glance.

A more compelling explanation of this behavior is that individuals bring their personal values to bear even when reputational considerations are absent, and are more or less inclined to behave in socially acceptable and morally approved ways, even when there is no material gain to be had by doing so. People often do what they do, quite simply because they believe it is the right thing to do.

One example of this propensity is strong reciprocity (Gintis 2000), accord-

ing to which individuals are predisposed to cooperated in a social dilemma and to punish free-riders, even at net personal cost. Another example is respect for character virtues such as honesty and trustworthiness, to which individuals conform not out of regard for others, but because virtuous behavior is its own reward (Gneezy 2005).

2 Rationality Does Not Imply Backward Induction

Suppose Alice and Bob play the Prisoner’s Dilemma, one stage of which is shown to the right, 100 times. Common sense tells us that players will cooperate for at least 95 rounds, and this is indeed supported by experimental evidence (Andreoni and Miller 1993). However, a backward induction argument indicates that players will defect on the very first round. To see this, note that the players will surely defect on round 100. But then, nothing they do on round 99 can help prolong the game, so they will both defect on round 99. Repeating this argument 99 times, we see that they will both defect on round 1.

	C	D
C	3,3	0,4
D	4,0	1,1

Given that common sense dictates cooperating for many rounds, and given that the players’ payoff is dramatically improved if they follow their common sense intuitions, in what sense is it “rational” to defect on every round? It is rational, as Robert Aumann showed in a famous paper (Aumann 1995), in the sense that common knowledge of rationality (CKR) implies backward induction. We define an agent who always chooses a best response as “rational.” We say CKR holds if each player knows the others are rational, each knows that each knows the others are rational, and so on, recursively, for all levels of mutual knowledge of rationality. Aumann’s theorem says that in an extensive form game with a unique subgame perfect Nash equilibrium, the only nodes on the game tree at which CKR can hold are along the backward induction path.

Prior to Aumann’s proof, many prominent game theorists had argued that rationality does not imply backward induction (Binmore 1987, Bicchieri 1989, Pettit and Sugden 1989, Basu 1990, Reny 1993), and many of these authors have been unpersuaded by Aumann’s proof, maintaining that it is inconsistent to assume CKR off the backward induction path. This critique, however, is incorrect. The backward induction argument is simply a classic example of *reductio ad absurdum*: assume a proposition and then show that this leads to a contradiction, proving that the proposition is false. Moreover, in the case of the repeated prisoner’s dilemma, the backward induction argument at a particular stage of the game never even assumes that we are off the backward induction path.

Neither Aumann nor anyone else has shown that rationality implies backward induction. Rather, Aumann has show that CKR implies backward induction. Economists

have generally presumed that in a world of rational agents, CKR is merely a universal recognition of this state of affairs. However, no one has ever proved CKR starting with plausible assumptions concerning inter-subjective knowledge. I claim in *Bounds of Reason* that CKR leads to logical paradoxes. Indeed, I present a version of CKR, which I term *common knowledge of logicality*, and claim that this concept is generally false, even when all agents are perfectly logical. Consider the following example.

Let us say an agent is *logical* in making inferences concerning a set of propositions if the agent rules out all statements that are inconsistent with this set. We then define *common knowledge of logicality* (CKL) for a set $i = 1, \dots, n$ of agents as follows. For any set of agents $i_1, \dots, i_k \in [1, \dots, n]$, i_1 knows that i_2 knows that \dots knows that i_{k-1} knows that i_k is logical.

Father William has \$690,000 to leave to his children, Alice and Bob, who do not know the size of his estate. He decides to give one child \$340,000 and the other \$350,000, each with probability 1/2. However, he does not want one child to feel slighted by getting a smaller amount, at least during his lifetime. So, he tells his children: “I will randomly pick two numbers, without replacement, from a set $S \subseteq [1, \dots, 100]$, assign to each of you randomly one of these numbers, and give you an inheritance equal to \$10,000 times the number you have been assigned. Knowing the number assigned to you will not allow you to conclude for sure whether you will inherit more or less than your sibling.” Father William, confident of the truth of his statement, which we take to be common knowledge for all three individuals, sets $S = \{34, 35\}$.

Alice ponders this situation, reasoning as follows, assuming CKL. Father knows that if $1 \in S$ or $100 \in S$, then there is a positive probability one of these numbers will be chosen and assigned to me, in which case I would be certain of the relative position of my inheritance. Alice knows that Father William knows she is logical, so she knows that $1 \notin S$ and $100 \notin S$. But, Alice reasons that her father knows that she knows that he knows she is logical, so she concludes that Father William knows that he cannot include 2 or 99 in S . But, Alice knows this as well, by CKL, so she reasons that Father William cannot include 3 or 98 in S . Completing this recursive argument, Alice concludes that S must be empty.

However, Father William gave one child the number 34, and the other 35, neither child knowing for sure which has higher number. Thus, Father William’s original assertion was true, and Alice’s reasoning was faulty. But, Alice’s only non-trivial assumption was CKL. We conclude that *common knowledge of logicality cannot be posited* in this context. CKL fails when the father included 35 in S , because this behavior is precluded by CKL.

CKL appears *prima facie* to be an innocuous extension of logicality, and indeed is not usually even mentioned in such problems, but in fact it leads to faulty

reasoning and must be rejected. In this regard, CKL is much like CKR, which also appears to be an innocuous extension of rationality, but in fact is often counterindicated.

It is not then surprising, as Vernon Smith stresses, that experimental subjects do not use backward induction. A better treatment of the 100 stage repeated prisoner's dilemma follows from simply assuming that each player believes the other will cooperate up to a certain round and defect thereafter, and has a Bayesian prior over the round on which his partner will first defect.

Pursuing this point, suppose Alice conjectures Bob will cooperate up to round k , and then defect thereafter, with probability g_k . Then, Alice will choose a round m to defect that maximizes the expression

$$\pi_m = \sum_{i=1}^{m-1} 3(i-1)g_i + (3(m-1)+1)g_m + (3(m-1)+4)(1-G_m), \quad (1)$$

where $G_m = g_1 + \dots + g_m$. The first term in this expression represents the payoff if Bob defects first, the second if Alice and Bob defect simultaneously, and the final term if Alice defects first. In many cases, maximizing this expression will suggest cooperating for many rounds for all plausible probability distributions. For instance, suppose g_k is uniformly distributed on the rounds $m = 1, \dots, 99$. Then, the reader can check by using (1) that it is a best response to cooperate up to round 98. Indeed, suppose Alice expects Bob to defect on round 1 with probability 0.95, and otherwise defect with equal probability on any round from 2 to 99. Then it is still optimal to defect on round 98. Clearly the backward induction assumption is not plausible unless you think your opponent is highly likely to be an obdurate backward inductor.

The reasoning dilemma begins if I then say to myself "My partner is just as capable as I of reasoning as above, so he will also cooperate at least up to round 98. Thus, I should set $m = 97$. But, of course my partner also knows this, so he will surely defect on round 96, in which case I should surely defect on round 95." This sort of self-contradictory reasoning shows that there is something faulty in the way we have set up the problem. If the $\{g_k\}$ distribution is reasonable, then I should use it. It is self-contradictory to use this distribution to show that it is the wrong distribution to use. But, my rational partner will know this as well, and I suspect he will revert to the first level of analysis, which says to cooperate at least up to round 95. Thus we two rational folks will cooperate for many rounds in this game rather than play the Nash equilibrium.

Suppose, however, that it is common knowledge that both I and my partner have the *same Bayesian prior* concerning when the other will defect. This is some-

times called *Harsanyi consistency* (Harsanyi 1967). Then, it is obvious that we will both defect on our first opportunity, because the backward induction conclusion now follows from a strictly Bayesian argument: the only prior that is compatible with common knowledge of common priors is defection on round one. However, there is no plausible reason for us to assume Harsanyi consistency in this case.

This argument reinforces our assertion that *there is nothing compelling about CKR*. Classical game theorists commonly argue that rationality *requires* that rational agents use backward induction, but in the absence of CKR, this is simply not the case.

3 Social Norms and Rational Action

The naive notion promoted in the textbooks, and dutifully affirmed by virtually every professional economist, is that rational agents play Nash equilibria. The repeated prisoner's dilemma presented in Section 2 shows that this is not the case, even in two player games with a unique Nash equilibrium, and where this equilibrium uses only pure strategies. If there are more players, if there are multiple equilibria, as is the general case in the sorts of repeated games for which some version of the Folk Theorem holds, or in principal agent models, or in signaling models, the presumption that the rationality assumption implies that agents play Nash equilibria is simply untenable. Of course, this fact is widely known, but there appears to be a "professional blindness" that bids us ignore the obvious.

Perhaps the most egregious, yet ubiquitous, example of ignoring the questionable status of the Nash equilibrium is that of mixed strategy Nash equilibria. For instance, suppose that Alice and Bonnie can each bid an integral number of dollars. If the sum of their bids is less than or equal to \$101, each receives her bid. If the total is exceeded, they each get zero. All symmetric equilibria have the form $\sigma = ps_x + (1-p)s_y$, where $x + y = 101$, $p \in (0, 1)$, and $x = py$, with expected payoff x for each player. In the most efficient equilibrium, each player bids \$50 with probability $p = 50/51$ and \$51 with probability $p = 1/51$.

But, if Alice plays the latter mixed strategy, then Bonnie's payoff to bidding \$50 equals her payoff to bidding \$51, so she has no rational incentive to play the mixed strategy. Moreover, Alice knows this, so she has no rational incentive to play any particular strategy. Thus, while it is intuitively plausible that they players would choose between bidding \$50 and \$51, and would choose the former with a much higher probability than the latter, this is certainly not implied by the rationality assumption.

Despite their apparent reticence to communicate this embarrassing truth to students, game theorists recognized that rational agents have no incentive to play

strictly mixed strategy Nash equilibria many years ago. One attempt to repair this situation was Harsanyi (1973), whose analysis was based on the observation that games with strictly mixed strategy equilibria are the limit of games with slightly perturbed payoffs that have pure strategy equilibria, and in the perturbed games, the justification of Nash behavior is less problematic. However, Harsanyi's approach does not apply to games with any complexity, including repeated games and principal-agent interactions Bhaskar (2000). The status of mixed strategy equilibria is restored in evolutionary game theory, because every equilibrium of an evolutionary dynamical system is a Nash equilibrium of the underlying stage game (Gintis 2009b, Ch. 11), but this fact does not help understand the relationship between rationality and Nash equilibrium.

In fact, the Nash equilibrium is *not* the equilibrium concept most naturally associated with rational choice. Robert Aumann (1987) has shown that the *correlated* equilibrium is the equilibrium criterion most worthy of this position. The concept of a correlated equilibrium of a game \mathcal{G} is straightforward. We add a new player to the game whom I will call the *choreographer* (more prosaically known as the 'correlating device') and a probability space (Σ, \tilde{p}) where Σ is a finite set and \tilde{p} is a probability distribution over Σ , which we call the *state space*. We assume also that there is a function $f: \Sigma \rightarrow S$, where S is the set of strategy profiles for the game \mathcal{G} . In effect, in state $\sigma \in \Sigma$, which occurs with probability $\tilde{p}(\sigma)$, the choreographer issues a directive $f_i(\sigma) \in S_i$ to each player $i = 1, \dots, n$ in the game, where S_i is player i 's pure strategy set. Note that $f_i(\sigma)$ may be correlated with $f_j(\sigma)$, so the choreographer can issue statistically correlated directives. For example, the system of red and green lights at a traffic intersection may be the choreographer, which simultaneously directs traffic in one direction to go (green) and in the other to stop (red). We say this configuration is a *correlated equilibrium* if it is a best response for each player to obey the choreographer's directive, providing all other players are likewise obeying.

To state Aumann's theorem, I must first define an *epistemic game*, which is a game \mathcal{G} , plus a set of possible states Ω . A state $\omega \in \Omega$ specifies, among other aspects of the game, the strategy profile $s(\omega)$ used in the game when the actual state is ω . In every state ω , each player knows only that the state is in some set of *possible* states $\mathbf{P}_i\omega \subset \Omega$. For instance, players may know their own moves, and perhaps their own types or other personal characteristics, but may not know other player's moves or types. Finally, each player has a *subjective prior* $p_i(\cdot; \mathbf{P}_i\omega)$ over Ω that is a function of the current state ω , but is the same for all states $\mathbf{P}_i\omega$ that i considers possible at ω . This subjective prior, $p_i(\cdot; \mathbf{P}_i\omega)$, represents precisely the player's beliefs concerning the state of the game, including the choices of the other players, when the actual state is ω .

Since each state ω in epistemic game \mathcal{G} specifies the players' pure strategy

choices $\mathbf{s}(\omega) = (\mathbf{s}_1(\omega), \dots, \mathbf{s}_n(\omega)) \in S$, the players' subjective priors must specify their beliefs $\phi_1^\omega, \dots, \phi_n^\omega$ concerning the choices of the other players. We call ϕ_i^ω i 's *conjecture* concerning the behavior of the other players at ω . A player i is deemed *rational* at ω if $\mathbf{s}_i(\omega)$ maximizes $\pi_i(s_i, \phi_i^\omega)$, where

$$\pi_i(s_i, \phi_i^\omega) =_{\text{def}} \sum_{s_{-i} \in S_{-i}} \phi_i^\omega(s_{-i}) \pi_i(s_i, s_{-i}), \quad (2)$$

where s_{-i} is a strategy profile of players other than i , S_{-i} is the set of all such strategy profiles, and $\pi_i(s_i, s_{-i})$ is the payoff to player i who chooses s_i when the other players choose s_{-i} .

We say the players $i = 1, \dots, n$ in an epistemic game have a *common prior* $p(\cdot)$ over Ω if there, for every state $\omega \in \Omega$, and every $i = 1, \dots, n$, $p_i(\cdot; \mathbf{P}_i\omega) = p(\cdot | \mathbf{P}_i\omega)$; i.e., each player's subjective prior is the conditional probability of the common prior, conditioned on i 's particular information $\mathbf{P}_i\omega$ at ω . We then have

Theorem 1. If the players in epistemic game \mathcal{G} are rational and have a common prior, then there is a correlated equilibrium in which each player is directed to carry out the same actions as in \mathcal{G} with the same probabilities.

The proof of this theorem is very simple, and consists basically of identifying the probability space Σ of the correlated equilibrium with the state space Ω of \mathcal{G} , and the probability distribution \tilde{p} with the common prior $p(\cdot)$.

This theorem suggests a direct relationship between game theory and the rational actor on the one hand, and the psycho-social theory of norms on the other. The common prior assumption, key to the association between Bayesian rationality and correlated equilibrium, is socially instantiated by a *common culture*, which all individuals in a society share (at least in equilibrium), and which leads them to coordinate their behaviors appropriately. Moreover, the choreographer of the correlated equilibrium corresponds to the *social norm*, which prescribes a particular behavior for each individual, according to the particular social roles they occupy in society.

It is important to note that this theorem holds even for self-regarding agents, which appears to imply that social norms could be effective in coordinating social activity even in the absence of a moral commitment to social cooperation. This may indeed be the case in some situations, but probably not in most. First, there may be several behaviors that have equal payoff to that suggested by the social norm for a particular individual, so a personal commitment to role performance may be required to induce individuals to play their assigned social roles. Second, individuals may have personal payoffs to taking certain actions that are unknown to the choreographer, and would lead amoral self-regarding agents to violate the

social norm's directive for their behavior. For instance, a police officer may be inclined to take bribes in return for overlooking criminal behavior, or a teacher may be inclined to favor a student of one ethnic group over another of a different background, thus ignoring the norms associated with their social roles. However, if the commitment to the ethic of norm compliance is sufficiently great, such preferences will not induce players to violate the duties associated with their roles.

The psycho-social theory of norms is a formal representation of Vernon Smith's suggested explanation of behavior in anonymous one shots. Smith says

Why should a real person see no continuation value across stage games with different but culturally more or less similar strangers? Can we ignore the fact that each person shares cultural elements of commonality with the history of others? . . . Is not culture about multilateral human sociality? These empirical extra theoretical questions critically affect how we interpret single play observations.

However, these are not "empirical extra theoretical questions." Rather they strike at the very heart of post methodological individualism social theory. It is the essence of human sociality that individuals generally behave ethically because it is the right thing to do, in addition to any motivation they might have due to long-term reputation effects or the possibility of being rewarded or punished for their actions. Smith call these questions "empirical" and "extra theoretical," but their epistemological status is just as broadly observed and theoretically grounded as any aspect of human choice behavior.

4 Preferences are Functions of Social Context

When individuals interact, each uses the social cues attendant to the interaction to assess the social norms appropriate to the interaction. However rare or unusual the context, each individual will generally have identified the situation with a customary interaction to which standard roles attach, and with which standard norms are associated. In a given society, most individuals will make the same assessment, especially if the context is relatively standard. Therefore, social cues will determine the expectations of the interacting individuals, leading to *common priors*. The social norms attached to the standard context will serve as a correlating device, and the agents will play a correlated equilibrium.

One obvious implication of the above scenario is that social cues determine, or at least strongly influence, the expectations players have of one another, and hence of their *beliefs*. Less obvious, but equally important, to the extent that individuals internalize the norms associated with the norms governing the interaction, they

will alter their preference ordering accordingly. Thus, preferences themselves are context-specific.

The dependence of both beliefs and preferences on social context explain in part the difficulty of modeling behavior in the laboratory.

The following experiment illustrates the fact that preferences are a function of social context. Dana, Cain and Dawes (2006) recruited 80 Carnegie-Mellon University undergraduate subjects who were divided into 40 pairs to play the dictator game, one member of each pair being randomly assigned Dictator, the other Receiver. Dictators were given \$10, and asked to indicate how many dollars each wanted to give the Receiver, but Receivers were not informed they were playing a Dictator Game. After making their choices, but before informing Receivers of the game, Dictators were presented with the option of accepting \$9 rather than playing the game. They were told that if a Dictator took this option, the Receiver would never find out that the game was a possibility, and would go home with their showup fee alone.

Eleven of the 40 Dictators took this exit option, including two who had chosen to keep all of the \$10 in the Dictator Game. Indeed, 46% of the Dictators who had chosen to give a positive amount to the Receiver took the exit option, in which the Receiver gets nothing. This behavior is not compatible with the concept of immutable preferences for a division of the \$10 between Dictator and Receiver, because individuals who would have given their Receiver a positive amount in the Dictator Game instead give them nothing by avoiding playing the game, and individuals who would have kept the whole \$10 in the Dictator Game are willing to take a \$1 loss not to have to play the game.

To rule out other possible explanations of this behavior, the authors executed a second study in which the Dictator was told that the Receiver would never find out that a Dictator Game had been played. Thus, if the Dictator gave \$5 to the Recipient, the latter would be given the \$5 along with the showup fee (with words to the effect "By the way, we are including an extra \$5"), but would given no reason why. In this new study, only one of 24 Dictators chose to take the \$9 exit option. Note that in this new situation, the same social situation between Dictator and Receiver obtains both in the Dictator Game and the exit option. Hence, there is no difference in the norms applying to the two options, and it does not make sense to forfeit \$1 simply to have the game not called a Dictator Game.

The most plausible interpretation of these results is that many subjects felt obliged to behave according to certain norms playing the Dictator Game, or violated these norms in an uncomfortable way, and were willing to pay simply not to be in a situation subject to these norms.

5 Conclusion

Experimental economics has vastly increased our knowledge of basic human behavior. In so doing, it has strengthened our appreciation for game theory and the rational actor model, because experimental methodology is firmly grounded in these two analytical constructs. On the other hand, experimental economics has validated the critics of the rational actor model by demonstrating that both beliefs and preferences are functions of social context, preferences take the individual's current state as a parameter, and maximization often entails using decision-making heuristics as a means of conserving information costs.

It would have been nice if strategic interaction could be explained by charting the logical implications of juxtaposing a number of Bayesian rational actors, as contemporary game theorists have attempted do. But, it cannot be done. Methodological individualism is, for better or worse, wrong.¹ It is wrong because our species developed by imbuing its members with a deep substrate of sociality (Boyd and Richerson 1985, Brown 1991). Experimental economics has shown us that the challenge is to model this substrate and chart its interaction with self-regarding objectives. This is the legacy of experimental economics.

REFERENCES

- Andreoni, James and John H. Miller, "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," *Economic Journal* 103 (May 1993):570–585.
- Aumann, Robert J., "Correlated Equilibrium and an Expression of Bayesian Rationality," *Econometrica* 55 (1987):1–18.
- , "Backward Induction and Common Knowledge of Rationality," *Games and Economic Behavior* 8 (1995):6–19.
- Basu, Kaushik, "On the Non-Existence of a Rationality Definition for Extensive Games," *International journal of Game Theory* 19 (1990):33–44.
- Bhaskar, V., "The Robustness of Repeated Game Equilibria to Incomplete Payoff Information," 2000. University of Essex.
- Bicchieri, Cristina, "Self-Refuting Theories of Strategic Interaction: A Paradox of Common Knowledge," *Erkenntnis* 30 (1989):69–85.
- , *The Grammar of Society: The Nature and Dynamics of Social Norms* (Cambridge: Cambridge University Press, 2006).
- Binmore, Kenneth G., "Modeling Rational Players: I," *Economics and Philosophy* 3 (1987):179–214.

¹For better I believe, for were it correct, we would all be no better than sociopaths.

- , *Natural Justice* (Oxford: Oxford University Press, 2005).
- Boyd, Robert and Peter J. Richerson, *Culture and the Evolutionary Process* (Chicago: University of Chicago Press, 1985).
- Brown, Donald E., *Human Universals* (New York: McGraw-Hill, 1991).
- Camerer, Colin, *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton, NJ: Princeton University Press, 2003).
- Dana, Justin, Daylian M. Cain, and Robyn M. Dawes, “What You Don’t Know Won’t Hurt Me: Costly (but quiet) Exit in Dictator Games,” *Organizational Behavior and Human Decision Processes* 100 (2006):193–201.
- Fudenberg, Drew and Jean Tirole, *Game Theory* (Cambridge, MA: MIT Press, 1991).
- Gigerenzer, Gerd and P. M. Todd, *Simple Heuristics that Make us Smart* (New York: Oxford University Press, 1999).
- Gintis, Herbert, “Strong Reciprocity and Human Sociality,” *Journal of Theoretical Biology* 206 (2000):169–179.
- , *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences* (Princeton, NJ: Princeton University Press, 2009).
- , *Game Theory Evolving* (Princeton, NJ: Princeton University Press, 2009). Second Edition.
- Gneezy, Uri, “Deception: The Role of Consequences,” *American Economic Review* 95,1 (March 2005):384–394.
- Harsanyi, John C., “Games with Incomplete Information Played by Bayesian Players, Parts I, II, and III,” *Behavioral Science* 14 (1967):159–182, 320–334, 486–502.
- , “Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points,” *International Journal of Game Theory* 2 (1973):1–23.
- Holmström, Bengt, “Moral Hazard in Teams,” *Bell Journal of Economics* 7 (1982):324–340.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky, *Judgment under Uncertainty: Heuristics and Biases* (Cambridge, UK: Cambridge University Press, 1982).
- Kreps, David M., *A Course in Microeconomic Theory* (Princeton, NJ: Princeton University Press, 1990).
- , Paul Milgrom, John Roberts, and Robert Wilson, “Rational Cooperation in the Finitely Repeated Prisoner’s Dilemma,” *Journal of Economic Theory* 27 (1982):245–252.
- Lewis, David, *Conventions: A Philosophical Study* (Cambridge, MA: Harvard University Press, 1969).

- Linton, Ralph, *The Study of Man* (New York: Appleton-Century-Crofts, 1936).
- Luce, R. Duncan and Howard Raiffa, *Games and Decisions* (New York: John Wiley, 1957).
- Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green, *Microeconomic Theory* (New York: Oxford University Press, 1995).
- Milgrom, Paul R. and Robert J. Weber, "A Theory of Auctions and Competitive Bidding," *Econometrica* 50 (September 1982):1089–1121.
- Myerson, Roger B., *Game Theory: Analysis of Conflict* (Cambridge, MA: Harvard University Press, 1991).
- Osborne, Martin J. and Ariel Rubinstein, *A Course in Game Theory* (Cambridge, MA: MIT Press, 1994).
- Parsons, Talcott, *Sociological Theory and Modern Society* (New York: Free Press, 1967).
- Pettit, Philip and Robert Sugden, "The Backward Induction Paradox," *The Journal of Philosophy* 86,4 (1989):169–182.
- Polanyi, Michael, *The Tacit Dimension* (New York: Doubleday & Co., 1966).
- Rasmusen, Eric, *Games and Information: An Introduction to Game Theory* (Cambridge, UK: Blackwell Scientific, 1989).
- Reny, Philip J., "Common Belief and the Theory of Games with Perfect Information," *Journal of Economic Theory* 59 (1993):257–274.
- Rubinstein, Ariel, "Perfect Equilibrium in a Bargaining Model," *Econometrica* 50 (1982):97–109.
- Smith, Vernon, "Theory and Experiment: What are the Questions?," *Journal of Economic Behavior and Organization* ??,?? (?? 2009):??–??.
- Tirole, Jean, *The Theory of Industrial Organization* (Cambridge, MA: MIT Press, 1988).