Journal of Economic Behavior & Organization 73 (2010) 34–40

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Discussion

Towards a renaissance of economic theory☆

Herbert Gintis

Santa Fe Institute and Central European University, 15 Forbes Avenue, Northampton, MA 01060, United States

ARTICLE INFO

Article history:
Received 22 September 2008
Accepted 22 September 2008

In his penetrating essay, "Theory and Experiment," Vernon Smith observes that subject behavior often diverges from game-theoretic predictions and asserts that minor fudging with theory is unlikely to solve our problems because "circumstances unknown to us" are responsible for the explanatory failures of neoclassical economic theory. I deal precisely with these issues in *The Bounds of Reason* (2009a). My remarks here draw upon several themes from this book relevant to specifying a refocused 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, beliefs (called "subjective priors"), linking choices to the relative probability of alternative outcomes, and they maximize their expected payoffs given these preferences and beliefs, subject to whatever material, informational, and other constraints they face. Many observers consider the experimental work of Daniel Kahneman and others (Kahneman et al., 1982; Gigerenzer and Todd, 1999) as destructive of the rational actor model. In fact, however, their research has considerably strengthened the rational actor model's explanatory power, albeit at the expense of increasing its complexity (Gintis, 2009a, Chs. 1 and 12). Most important, these studies have shown the value of modeling using heuristics that conserve on information processing and of including the current state (physiological, temporal and ownership) of the individual as an argument of the preference ordering.

In the 1980s, 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 a centerpiece of microeconomic theory alongside the traditional Walrasian model of general market equilibrium. It is important to note, however, that this exalted position is quite recent. The textbooks that expound the received wisdom of today's practicing economists were written in the early years of the game theory revival, between 1988 and 1995. These textbooks include misconceptions promulgated in the 1980s 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.

I here 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, despite their analytical power, game theory and the rational actor model provide only part of the theoretical toolkit needed to explain human strategic interaction. Moreover, because individuals bring much of the complexity of society into 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 following material is adapted from Herbert Gintis. *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.

E-mail address: hgintis@comcast.net.

0167-2681/$ – see front matter
doi:10.1016/j.jebo.2008.09.012

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.

These changes present enormous challenges to economic theory. Each of the above suggested principles can be expressed analytically and validated experimentally, but together they specify an axiomatic system for which analytically closed solutions cannot be attained without significant theoretical innovation. It is not implausible to foresee an impending renaissance of economic theory.

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 1980s, the year 1982 alone seeing the publication of Rubinstein's famous game-theoretic model of bargaining, Milgrom and Weber's theory of auctions, Holmström's game-theoretic microfoundations of organizations, and the "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 et al. (1995). The Nobel prize in economics was awarded to a game theorist in the years 1994, 1996, 2001, 2005, and 2007.

The early textbooks promoted 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. In fact, this principle is not even approximately true even in extremely well-behaved cases, such as when such an equilibrium is the unique equilibrium of the game. The conditions under which rational agents choose subgame perfect Nash equilibria is important, but unsolved problem.

Though widely known, this message has not filtered through to the textbooks and hence is generally ignored by working economists. Not surprisingly, when experiments show 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 (this issue), who wisely rejects this facile answer and bids that we search for something more satisfying. It is important to note that contemporary game theory is not falsified by such experimental findings; rather, in most cases, game theory is too weak to predict anything at all.

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) that 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.

However, 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. 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 appropriate 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. Vernon Smith's Assumption 6 (Context does not matter, only the underlying abstract game structure) is incorrect precisely for this reason.

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 neoclassical model, rational actors are self-regarding unless expressing social preferences allows them to build reputations for cooperation in the future. An extensive body of experimental evidence supports the fact that individuals exhibit other-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 attributing subject behavior that violates neoclassical precepts to faulty cognition, does so in this case, saying "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, other-regarding behavior in one-shot interactions is a daily commonplace. 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 other-regarding 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), according to which individuals are predisposed to cooperate 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 many rounds, and this is supported by experimental evidence (Andreoni and Miller, 1993). However, a backward induction argument indicates that players will defect on the first and every succeeding round.

	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? As Robert Aumann showed in a famous paper (Aumann, 1995), not rationality, but rather 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, several 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 rationality off the backward induction path. This critique, however, is incorrect. The backward induction argument is simply a classic example of *reductio ad adsurdum*: 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. In fact, Aumann's theorem is completely consistent with the critics' assertion that rationality does not imply backward induction because there is no known set of plausible epistemic conditions under which rationality implies CKR.

Economists have generally presumed that in a world of rational agents, CKR is merely a universal recognition of this state of affairs. However, this is not the case. To clarify this point, let us define an *epistemic game* to be a game 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* $\mathrm{p}_i(\cdot\,;\mathbf{P}_i\omega)$ 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.

Consider any epistemic game with state space Ω that has a perfect information extensive form and a unique subgame perfect Nash equilibrium $\mathbf{s}^*$. The priors $p_i(\cdot\,|\omega)$ for $\omega \in \Omega$ fully determine the probability each player places on the occurrence of $\mathbf{s}^*$ namely, $p_i([\mathbf{s} = \mathbf{s}^*]|\omega)$. This is surely zero unless $\mathbf{s}_i(\omega) = \mathbf{s}^*$. Moreover, we must have $p_i([\mathbf{s} = \mathbf{s}^*]|\omega) = 1$ if $\omega \in$ CKR, which is a restriction on subjective priors that has absolutely no justification in general, although it can be justified in certain cases (e.g., the one- or two-round Prisoner's Dilemma).

It might be suggested that a plausible strategy selection mechanism epistemically justified by some principle other than CKR might succeed in selecting out the subgame perfect equilibrium—for instance, extensive form rationality as proposed by Pearce (1984) and Battigalli (1997). However, this selection mechanism is not epistemically grounded at all. There are alternative, epistemically grounded selection mechanisms for extensive form games, such as Fudenberg et al. (1988), Borgers (1994), and Ben-Porath (1997), but these mechanisms do not justify backward induction.

It is thus not surprising that, as Vernon Smith stresses, 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

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

where $G_m = g_1 + \ldots + g_m$. 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 $m = 1, \ldots, 99$. 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.

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 sometimes 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 us to understand the relationship between rationality and Nash equilibrium.

In fact, the Nash equilibrium is *not* the equilibrium concept most naturally associated with rational choice. 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 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 $(\Sigma, \tilde{p})$ 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: $\Sigma \to S$, where S is the set of strategy profiles for the game 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 $i = 1, \ldots, n$ 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, note that since each state ω in epistemic game G specifies the players' pure strategy choices $\mathbf{s}(\omega) = (\mathbf{s}_1(\omega), \ldots, \mathbf{s}_n(\omega)) \in S$, the players' subjective priors must specify their beliefs $\phi_1^\omega, \ldots, \phi_n^\omega$ 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

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

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 $i = 1, \ldots, n$ in an epistemic game have a *common prior* $p(\cdot)$ over Ω if there, for every state $\omega \in \Omega$, and every $i = 1, \ldots, n$, $p_i(\cdot; \mathbf{P}_i\omega) = p(\cdot|\mathbf{P}_i\omega)$; 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. *If the players in epistemic game 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 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 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 the individual occupies 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.

It is not hard to see, however, that these are deeply theoretical, not "empirical extra theoretical" questions. Indeed, widely observed ethical behavior dramatically reveals the standard game theory's error in embracing not just the self-regarding actor model, but the whole philosophy of methodological individualism. Individuals often behave ethically because it is the right thing to do, whether or not they will be rewarded in the future for their prosocial behavior. Moreover, individuals are predisposed to consider an action to be the right thing to do when it conforms to a legitimate social norm. An adequate theory of human behavior must successfully incorporate these facts.

4. Preferences are functions of social context

When individuals in a group interact, each member 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 this line of analysis is that social cues 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 governing the interaction, they will alter their preference orderings accordingly. Thus, in a deep sense preferences themselves are context-specific.

The following experiment illustrates the fact that preferences are a function of social context. Dana et al. (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 percent 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 undertook 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 be 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 demonstrated 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. These findings do not suggest that we abandon the rational actor model, but rather that we adopt a more complex version thereof.

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, contradicted by the evidence. Our species developed by imbuing its members with a deep, but not irrational, 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. The legacy of experimental economics may thus be a renaissance of economic theory.

References

Andreoni, J., Miller, J.H., 1993. Rational cooperation in the finitely repeated prisoner’s dilemma: experimental evidence. Economic Journal 103, 570–585.
Aumann, R.J., 1987. Correlated equilibrium and an expression of Bayesian rationality. Econometrica 55, 1–18.
Aumann, R.J., 1995. Backward induction and common knowledge of rationality. Games and Economic Behavior 8, 6–19.
Basu, K., 1990. On the non-existence of a rationality definition for extensive games. International Journal of Game Theory 19, 33–44.
Battigalli, P., 1997. On rationalizability in extensive form games. Journal of Economic Theory 74, 40–61.
Ben-Porath, E., 1997. Rationality Nash equilibrium and backward induction in perfect-information games. Review of Economic Studies 64, 23–46.
Bhaskar, V., 2000. The Robustness of Repeated Game Equilibria to Incomplete Payoff Information. University of Essex.
Bicchieri, C., 1989. Self-refuting theories of strategic interaction: a paradox of common knowledge. Erkenntniss 30, 69–85.
Bicchieri, C., 2006. The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press, Cambridge.
Binmore, K.G., 1987. Modeling rational players: I. Economics and Philosophy 3, 179–214.
Binmore, K.G., 2005. Natural Justice. Oxford University Press, Oxford.
Borgers, T., 1994. Weak dominance and approximate common knowledge. Journal of Economic Theory 64, 265–276.
Boyd, R., Richerson, P.J., 1985. Culture and the Evolutionary Process. University of Chicago Press, Chicago.
Brown, D.E., 1991. Human Universals. McGraw-Hill, New York.
Camerer, C., 2003. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton, NJ.
Dana, J., Cain, D.M., Dawes, R.M., 2006. What you don’t know won’t hurt me: costly (but quiet) exit in dictator games. Organizational Behavior and Human Decision Processes 100, 193–201.
Fudenberg, D., Kreps, D.M., Levine, D., 1988. On the robustness of equilibrium refinements. Journal of Economic Theory 44, 354–380.

Fudenberg, D., Tirole, J., 1991. Game Theory. MIT Press, Cambridge, MA.
Gigerenzer, G., Todd, P.M., 1999. Simple Heuristics That Make Us Smart. Oxford University Press, New York.
Gintis, H., 2000. Strong reciprocity and human sociality. Journal of Theoretical Biology 206, 169–179.
Gintis, H., 2009a. The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences. Princeton University Press, Princeton, NJ.
Gintis, H., 2009b. Game Theory Evolving, Second edition. Princeton University Press, Princeton, NJ.
Gneezy, U., 2005. Deception: the role of consequences. American Economic Review 95, 384–394.
Harsanyi, J.C., 1967. Games with incomplete information played by Bayesian players Parts i, ii, and iii. Behavioral Science 14, 159–182, 320–334, 486–502.
Harsanyi, J.C., 1973. Games with randomly disturbed payoffs: a new rationale for mixed-strategy equilibrium points. International Journal of Game Theory 2, 1–23.
Holmstrom, B., 1982. Moral hazard in teams. Bell Journal of Economics 7, 324–340.
Kahneman, D., Slovic, P., Tversky, A., 1982. Judgment under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge, UK.
Kreps, D.M., 1990. A Course in Microeconomic Theory. Princeton University Press, Princeton, NJ.
Kreps, D.M., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in the finitely repeated prisoner's dilemma. Journal of Economic Theory 27, 245–252.
Lewis, D., 1969. Conventions: A Philosophical Study. Harvard University Press, Cambridge, MA.
Linton, R., 1936. The Study of Man. Appleton-Century-Crofts, New York.
Luce, R.D., Raiffa, H., 1957. Games and Decisions. John Wiley, New York.
Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic Theory. Oxford University Press, New York.
Milgrom, P.R., Weber, R.J., 1982. A theory of auctions and competitive bidding. Econometrica 50, 1089–1121.
Myerson, R.B., 1991. Game Theory: Analysis of Conflict. Harvard University Press, Cambridge, MA.
Osborne, M.J., Rubinstein, A., 1994. A Course in Game Theory. MIT Press, Cambridge, MA.
Parsons, T., 1967. Sociological Theory and Modern Society. Free Press, New York.
Pearce, D., 1984. Rationalizable strategic behavior and the problem of perfection. Econometrica 52, 1029–1050.
Pettit, P., Sugden, R., 1989. The backward induction paradox. The Journal of Philosophy 86, 169–182.
Polanyi, M., 1966. The Tacit Dimension. Doubleday Co, New York.
Rasmusen, E., 1989. Games and Information: An Introduction to Game Theory. Blackwell Scientific, Cambridge, UK.
Reny, P.J., 1993. Common belief and the theory of games with perfect information. Journal of Economic Theory 59, 257–274.
Rubinstein, A., 1982. Perfect equilibrium in a bargaining model. Econometrica 50, 97–109.
Smith, V., this issue. Theory and experiment: what are the questions? Journal of Economic Behavior and Organization.
Tirole, J., 1988. The Theory of Industrial Organization. MIT Press, Cambridge, MA.