CHAPTER 13

Behavioral Game Theory: Predicting Human Behavior in Strategic Situations

COLIN F. CAMERER

1. Introduction

In strategic interactions, what one player does affects another player's payoff. Game theory is a mathematical language for describing strategic interactions and their likely outcomes (Fudenberg and Tirole 1991; Osborne and Rubinstein 1995). A "game" is a specification of the strategies each of several "players" have, the order in which players choose strategies, the information players have, and how players rate the desirability ("utility") of resulting outcomes. Game theory is flexible enough to be used at many levels of detail in a broad range of sciences. Players can be genes, people, groups, firms, or nation-states. Strategies can be genetically coded instincts, methods of bidding on Ebay, corporate practices for developing and introducing new products, a legal principle in a complex mass tort case, or wartime battle plans. Outcomes can be anything players value—prestige, food, control of Congress, sexual opportunity, corporate profits, a sense of justice, or captured territory.

Even without doing any mathematical analysis, game theory can be useful as a taxonomy that parses the strategic world (Aumann 1985). Analytical game theory goes further, deriving precise predictions about how players might behave by assuming that players maximize expected utility, plan ahead, and form beliefs about other players' likely moves (by assuming those players plan and maximize also).

While game theory is a powerful analytic engine, hundreds of experiments show that its predictions are systematically violated (Crawford 1997; Camerer 2003). Violations of any simple theory can be comfortably tolerated unless they point to an easy way to improve the theory. This chapter describes an emerging approach called "behavioral game theory," which generalizes analytical game theory to explain experimentally observed violations and respect bounds on human cognition. An analytical game theorist crossing a one-way street only looks one way before crossing the street (the only direction that rational drivers would come from); a behavioral game theorist looks both ways, anticipating possible mistakes. However, the goal is to establish regularity of these mistakes empirically, and tie theories of them to psychological and biological principles.

While the theory is inspired by laboratory regularity, it is aimed at practical questions like worker reactions to employment terms, evolution of internet market

C.F. Camerer, G. Loewenstein & M. Rabin (eds.), *Advances in Behavioral Economics*, (New York: Princeton University Press, 2004), 374-392.

institutions for centralized trading, animal behavior, and players "teaching" other players who learn what to expect (like firms intimidating competitors or building trust in strategic alliances, or diplomats threatening and cajoling). Behavioral game theory also exemplifies the potential from reunification of psychology and economics, which wandered apart from the 1920s until recently, as psychologists pursued empirical laboratory regularity and economists practiced using simple formal models to understand field data and evaluate policy.

Game theory has a clear paternity. After some important early contributions, its main features were introduced by John von Neumann and Oskar Morgenstern in 1944. Shortly thereafter, John Nash (1950) proposed a general solution to the problem of how rational players would play. Nash suggested that players adjust their strategies until they reached an "equilibrium" in which any unilateral adjustment was not beneficial (a fixed point in the mapping from strategies to the set of best-response strategies). In 1995, Nash shared a Nobel Prize with John Harsanyi and Reinhard Selten for their pioneering work on games played over time, and games in which players have private information about their motivations (Nash 1950; Selten 1975; Harsanyi 1967–68).

Game theory has been used in many social science applications, mostly in economics but increasingly in political science, biology, sociology, psychology, and anthropology. Analytical game theory has been used to model phenomena like price competition among firms and R&D investments in "patent races" (Tirole 1988), coordination when there are synergies between firms or in the macroeconomy (Cooper 1999), political candidates positioning themselves in an "issue space" to maximize votes (Shepsle and Boncheck 1997), divorces, bankruptcies, and strikes (Kennan and Wilson 1990), conflicts between "principals" and the "agents" whom they hire to work for them—such as managers and workers (Milgrom and Roberts 1992), and animal behavior (Maynard Smith 1982).

Game theory began as applied mathematics and spawned many intriguing puzzles, so there is much more theory than direct observation. Tests with field data have occurred in only a few areas—auctions for oil leases and airwave spectrum (Hendricks and Paarsch 1995; Laffont 1997; McAfee and McMillan 1996), incentive contracting (Prendergast 1999), industrial organization (Bresnahan and Reiss 1991), labor-management bargaining (Kennan and Wilson 1990), and matching of medical residents, sororities, and college bowl games (Roth and Peranson 1999). Field tests are problematic because predictions about equilibrium behavior often depend very sensitively on players' strategies, information, and payoffs, which are usually not observable. Experiments that control these details are therefore particularly helpful.

Hundreds of experiments have been conducted in recent years (Camerer 2003). Three elements of behavioral game theory which explain these experimental findings are: social utility functions; initial conditions (first-period play); and learning theories. Next I give examples of experimental findings illustrating each element. Colman (in press) gives a complementary perspective with more attention to philosophical difficulties and unsolved problems in modeling coordination.

2. Social Utility

A social utility function expresses how players feel about the outcome a game, including the payoffs other players receive. The default social utility function in economic theory is that people care only about their own outcomes. This simplification, while useful, leaves out forces like altruism, fairness, trust, vengeance, hatred, reciprocity, and spite.

A famous example of how social motives affect behavior is the prisoners' dilemma (PD), in which players are collectively better off if they all cooperate, but prefer to defect whether others cooperate or not. Contrary to self-interest, in the lab players cooperate in the PD about half the time, typically when they expect others to cooperate. Other evidence of social motives comes from simple games like ultimatum bargaining. In an ultimatum game a Proposer is endowed with a sum, often $10, and offers a share to another player, the Responder. If the Responder rejects the offer they both get nothing. While the ultimatum game is only a building block of more complex natural bargaining (corresponding to "11th hour" offers on the courthouse steps), it is a convenient tool to measure whether Responders will sacrifice their own earnings to punish others who self-servingly violate norms of fair treatment.

These experiments typically pair subjects together anonymously for one play of the game, to establish a benchmark of how strangers in temporary situations behave. Assuming self-interest, game theory predicts that Responders will accept any positive amount, and Proposers will anticipate this and offer very little. In fact, Responders typically reject offers of $2–3 half the time. Proposers seem to guess this and offer $4–5 (see figure 13.1) (Güth, Schmittberger, and Scharze 1982; Camerer and Thaler 1995). This result scales up to higher ($100) stakes with American college students (some rejected $30) and in low-income countries where modest payments equal 2–3 months' wages (Hoffman, McCabe, and Smith 1996). Norms and judgments of fairness can depend on context and culture. When Proposers earn the right to make the offer by winning at trivia, they feel entitled to offer less—and Responders accept less (Hoffman, McCabe, Shachat, and Smith 1994).

Dramatic new experiments show the effect of culture. Figure 13.2 shows data from Pittsburgh and a small Peruvian agricultural group, the Machiguenga (Henrich 2000). The Americans usually offer half (and, incidentally, often reject low offers). The Machiguenga offer much less (typically 15–25%) and rejected only one offer. It is ironic that the Machiguenga—one of the most culturally and economically primitive groups ever studied—come closest to the game theory prediction! Anthropologists have now studied ten other primitive cultures and found interesting variations in bargaining, which seem to be related to the degree of cooperation in economic activity (e.g., do men hunt collectively?) and degree of exposure to impersonal market trading (Henrich et al. 2000).

Ultimatum games tap negative reciprocity or vengeance. Other games reveal other motives. In dictator games, a Proposer simply dictates an allocation of

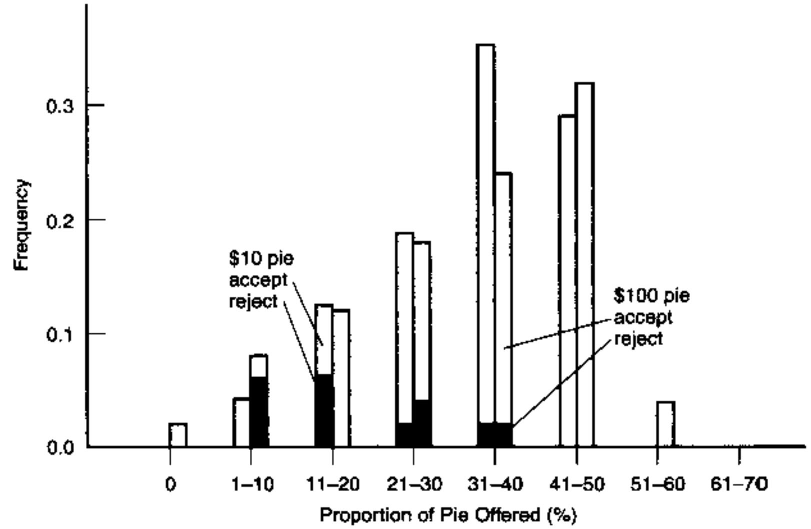

Figure 13.1 Offers and rejections in $10 and $100 ultimatum games.
Source: Hoffman et al. 1994.

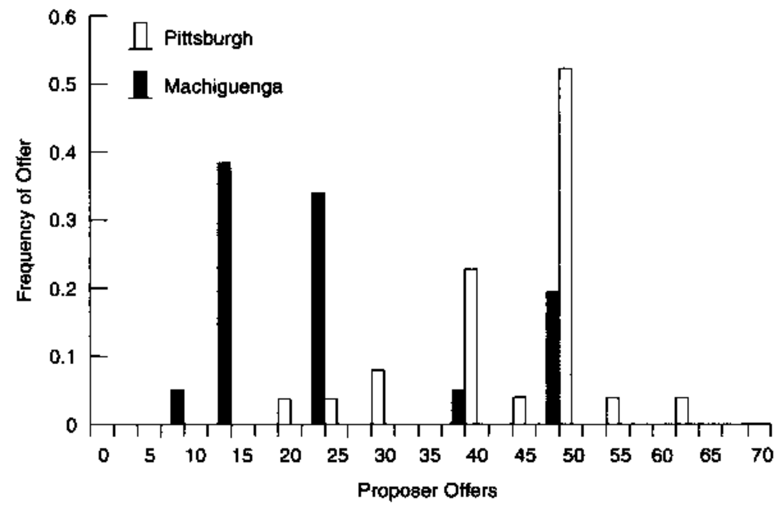

Figure 13.2 Proposer offers in ultimatum games are different for college students in Pittsburgh (26) and Machiguenga farmers in Peru (22). Offers are percentages offered to Responders. Mean, standard deviation, and sample size are .26, .14, 21 (Machiguenga) and .45, .10, 27 (Pittsburgh).

money and the Responder must accept it. In these games, Proposers offer less than in ultimatum games, about 15% of the stakes, but average offers vary widely with contextual labels and other variables (Camerer 2003).

In trust games, an Investor risks some of her endowment of money, which is tripled by the experimenter (representing a return on social investment) and handed to an anonymous Trustee. The Trustee pays back as much of the tripled sum as she likes to the Investor (perhaps nothing) and keeps the rest. Trust games are models of opportunities to gain from investment with no legal protection against theft by a business partner. Game theory predicts that self-interested Trustees will never pay back money; Investors will anticipate this and invest nothing. In fact, Investors typically risk about half their money, and Trustees pay back slightly less than was risked (Berg, Dickaut, and McCabe 1995), even when stakes are high (McKelvey and Palfrey 1992). Trustee payback is consistent with positive reciprocity.

A very important point is that competition has a strong effect in these games. If two or more Proposers make offers in an ultimatum game, and a single Responder accepts the highest offer, then the only equilibrium is for the Proposers to offer almost all the money to the Responder (the *opposite* of the prediction with one Proposer). In the lab this Proposer competition does occur rapidly: resulting in a very unfair allocation—almost no earnings for Proposers (Roth et al. 1991).

A good social utility theory could explain *with a single model* why Responders reject unfair offers, dictator Proposers give away money, Trustees repay trust, and why multiple Proposers compete to earn very little (and perhaps where such preferences came from; Gale, Binmore, and Samuelson 1995; Nowak, Page, and Sigmund 2000; Samuelson 2000).[1] In "inequality-aversion" theories, players prefer more money and also prefer that allocations be more equal. They will sacrifice some money to make outcomes more equal. In one such approach (Fehr and Schmidt 1999; cf. Bolton and Ockenfels 2000), player i's utility when x_k is the payoff to player k is $u_i(x_1, x_2, \ldots, x_n) = x_i + \Sigma_{k=1}^{n}\alpha(x_k - x_i)_0/(n-1) + \Sigma_{k=1}^{n}\beta(x_i - x_k)_0/(n-1)$ (where $(x)_0$ denotes the maximum of x and 0). The coefficients α and β represent the weight of envy and guilt, respectively. When these coefficients are zero, players are purely self-interested, so the standard model is a special case of this one. There are undoubtedly individual differences in these coefficients, with some degree of cross-game reliability, and they may be correlated with psychometric scales (e.g., Gunnthorsdottir, McCabe, and Smith 2002). Models like this can explain all the patterns mentioned above. Responders reject low offers to enforce equality. Dictator Proposers and Trustees create more equality by giving money to others. In the face of competition, letting another Proposer outbid you gives you no money *and* creates a multiple dose of envy (an empty-handed Proposer is envious of the Responder and of the Proposer whose bid was

[1] Social preferences are thought to have evolved in the ancestral past when humans lived in small groups. In such groups, collective gain from cooperation in the absence of property rights is enhanced by positive reciprocity. Negative reciprocity ensures that players get a share of joint outcomes. While these evolutionary explanations are surely part of the story, they do not naturally account for the strong influence of culture and contextual variables like entitlement and excuses.

TABLE 13.1
A Dictator Game with Unknown Recipient Payoffs

		Recipient Payoffs	
Dictator Choice	*Dictator Payoff*	*Sacrifice Game*	*No-sacrifice Game*
A	6	1	5
B	5	5	1

Source: Dana, Weber, Kuang, 2003.

accepted); the only way to get more money and less envy is to outbid the other Proposer. This also explains why fairness seems to play no prominent role in competitive double auctions: In those experiments, players usually do not know how much others earn (so the utility functions above don't apply); and even if they do, when there is competition sacrificing money does not improve equality so one might as well just maximize one's own payoff.

In reciprocity theories, player A forms a judgment about whether another player B has sacrificed to benefit (or harm) her (Rabin 1993). Player A likes to reciprocate, repaying kindness with kindness, and meanness with vengeance. This idea can also explain most of the results previously mentioned. (Blount 1995), and also explains the observed correlation between cooperation and expectations of cooperation by others in the PD.[2] Inequality-aversion and reciprocity theories differ because inequality-averse players care only about final allocations, while reciprocal players care about the events that led to the allocations (since they affect perceptions of kindness).

Recent evidence from Dana, Weber, and Kuang (2003) complicates the conclusion that simply modifying social utility is an adequate explanation. Their dictator game is illustrated in table 13.1. The dictator must choose row A or B. If she chooses A, she always earns 6; if she chooses B she earns 5. The payoffs to the recipient player depend on whether she is playing a sacrifice game or a no-sacrifice game. In the sacrifice game, the dictator can give up a point, picking B rather than A, to raise the recipient's payoff from 1 to 5 (and make the two players' payoffs equal). In the no-sacrifice game A Pareto-dominates (both players earn more than B). Not surprisingly, in the no-sacrifice game all the dictators choose B. In the sacrifice game 74% sacrificed by choosing B, consistent with theories above in which players will "spend" money to achieve equality.

These results are not surprising. In the more interesting condition, dictators were told that they were equally likely to be playing the sacrifice and no-sacrifice games, and they could choose whether or not to find out which game they were playing before making their choice. Players with social preferences for equality would want to know which game they are playing, since they would pick B and A in the two different games. But almost half the players chose *not* to find out, and

[2] Effectively, the PD becomes a game in which players are trying to coordinate their levels of emotion or reciprocity, and hence it has two pure equilibria rather than one.

85% of those players chose A. It appears that many players will sacrifice to help others if they "have to," but that they will also avoid finding out whether they are socially obliged to help. This pattern cannot be reconciled with social preference theories and requires more delicate concepts of rule-bound behavior (Rabin 1993) or self-identity and rationalization. Refusing to find out how one's behavior impacts others is important in some economic settings. For example, there is anecdotal evidence that many people who are at high risk for HIV infection refuse to get tested so that they can continue risky activities—which jeopardize others—without feeling that they are *knowingly* causing harm.

Regardless of which functional forms and complications prove most useful, social utility functions like these could be applied to explain charitable contribution, legal conflict and settlement, wage-setting, and wage dispersion within firms, strikes, divorces, wars, tax policy, and bequests by parents to siblings. Explaining these phenomena with a single parsimonious theory would be very useful.

3. Iterated Reasoning in First-Period Play and One-Shot Games

In many strategic situations players engage in the same game repeatedly. This raises two questions: How do they play the first time? How do they learn over time?

A theory of first-period play will be a statistical collage of ideas from decision theory and cognitive psychology. Some players choose randomly. Other players know they need to guess what others will do, and they "iterate" their reasoning by imagining what others will do, what others imagine others will do, and so forth. In game theory, this iterated reasoning is assumed to continue until a mutual best-response fixed-point is reached.

In the human mind, iterated reasoning surely halts after a small number of steps for several reasons. There will be evolutionary selection against high levels of iterated reasoning (Stahl 1993) if dedicating cortex to strategize against increasingly strategizing humans has an increasing marginal cost. "I think he thinks . . ." reasoning also taxes limited working memory (and cannot be stretched by chunking items). And overconfidence may lead players to think others have not thought as deeply as they have, braking the iteration (Costa-Gomes, Crawford, and Broseta 2001).

Most importantly, many experiments show that 0–2 steps of iterated reasoning are likely in the first period of play, even among analytically brilliant college students and Ph.D.'s (Nagel 1999; Ho, Camerer, and Weigelt 1998). An illustration is the "*p*-beauty contest." In this game, several players choose a number in the interval [0,100]. The average of the numbers is computed, and multiplied by p (say [2/3]). The player whose number is closest to [2/3] of the average wins a fixed prize. The game is called a *p*-beauty contest after a famous book by Keynes (1936, p. 34). He likened investing in stocks to a beauty contest in which players just wanted to guess who others thought was most beautiful (a metaphor that is particularly apt for tech-stock "bubbles").

The *p*-beauty contest is a good way to measure steps of iterated reasoning. Some people appear to choose randomly, or pick a favorite number, exhibiting "step-0" thinking. Players who think others choose randomly can guess that the average will be 50, so they should choose (2/3) of 50, or 33 (step-1 iteration). If a player thinks others think that way, she should choose 22. The numbers players choose reveal how the number of iterations of reasoning. Nash equilibrium requires mutual best-response, so it does not stop until it reaches a common number $x = (2/3)x$. So the Nash equilibrium is zero.

Figure 13.3 shows results from the $p = 2/3$ game played for $20, in four subject pools (Camerer 2003, chapter 5). Students exhibit 0–3 steps of iterated thinking, choosing numbers which average around 25. (Caltech students more than half of whom have median math SAT score of 800 choose lower numbers than other students, but rarely choose zero.) Ph.D.'s, portfolio managers, and a sample of successful CEOs behave much like students do. When the game is played ten times with the same players (who learn the average after each trial), numbers converge toward zero (see figure 13.4), a reminder that equilibrium concepts help predict where an adaptive process leads.

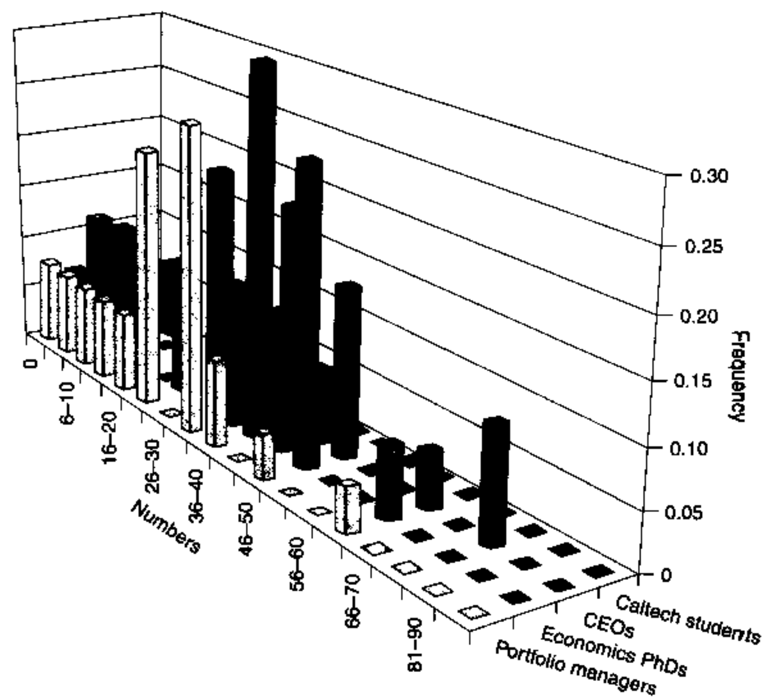

Figure 13.3 Number choices in *p*-beauty contest games by 4 subject pools (3).

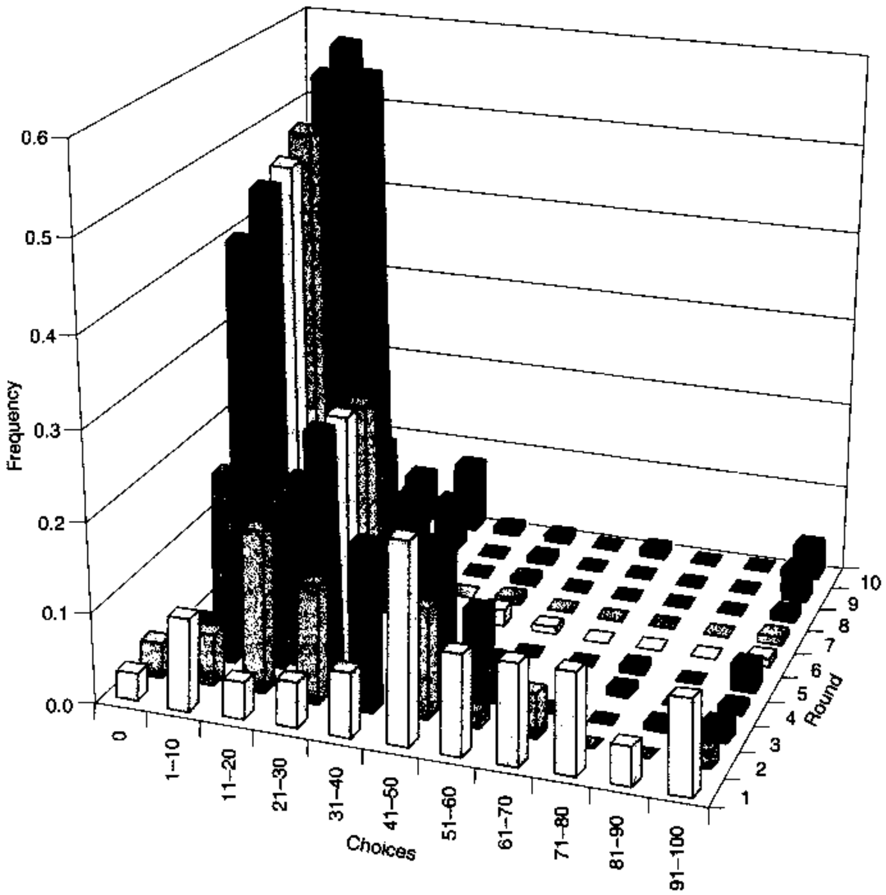

FIGURE 13.4 Number choices in p-beauty contest played 10 times (44).

Other games in which strategy choices correspond to steps of iterated thinking show similar regularity in reasoning levels (Camerer 2003). (An example is Bertrand competition, in which firms selling an identical product undercut each others' prices until they all sell at the marginal cost. Internet-based pricing of bestselling books and other commodities seems headed in this direction.) A natural model is one in which players use different levels of iterated reasoning. Statistical estimates suggest (using figure 13.4 data) 10–20% of the subjects using each of 0–3 steps. Camerer, Ho, and Chong (2003) use a one-parameter Poisson distribution to characterize the distribution of thinking steps in a "cognitive hierarchy" (CH). They estimate the average number of steps, τ, to be between 1–2 across almost a hundred games. Models of this type are more cognitively plausible, more descriptively accurate (of one-shot

experimental data) than equilibrium concepts, and they are also *more precise* than Nash equilibrium (once τ is specified), because the CH model predicts a specific statistical distribution even when there are multiple Nash equilibria.[3]

The CH model suggests why Nash equilibrium predicts surprisingly well in some classes of one-shot games, like those with mixed equilibria. Here's why: The best-responses by players who do increasingly many thinking steps tend to cycle among the strategies which are played with positive probability in the mixed equilibrium. So the model created a kind of endogeneous "purification" in which players at different thinking steps play pure strategies (they think they have "figured it out" and may not think of themselves as randomizing), but the mixture across those pure strategies can closely resemble the mixture predicted by Nash equilibrium.

Measuring steps of reasoning ignores the benefits and costs of thinking hard. Costs and benefits can be included by relaxing Nash equilibrium, so player i is assumed to form beliefs about the chance that other players s_{-i} will choose strategy k (denoted $P(s^k_{-i})$). Then i calculates an expected payoff for her own strategy J, denoted $\Sigma_k P(s^k_{-i})\pi_i(s^j_i, s^k_{-i})$ and chooses better responses more often than bad responses, according to a logit response rule, $P(s^j_i) = \exp(\lambda\Sigma_k P(s^k_{-i})\pi_i(s^j_i, s^k_{-i})/(\Sigma_j \exp(\lambda\Sigma_k P(s^k_{-i})\pi_i(s^j_i, s^k_{-i}))$. A "quantal response" equilibrium (QRE) exists when each player's beliefs about choice probabilities of others are consistent with the actual choice probabilities of others.

QRE is a competent one-parameter generalization of Nash equilibrium for fitting experimental data (McKelvey and Palfrey 1995, 1998; Georee and Holt 1999). It acknowledges that players will sometimes choose strategies that appear, to an observer, to sacrifice payoffs, but assumes that big mistakes are rarer than small ones. It also circumvents technical limits of Nash equilibrium and jibes with intuitions (and data) about many of the quirkier predictions of analytical game theory (Goeree and Holt 2001).[4]

Theories of statistical mixtures of iterated reasoning, or QRE, could predict initial reactions of consumers and voters to economic and policy changes better than equilibrium theories. Initial conditions are important because they can be influential in determining the direction and path of convergence after a change (particularly when there is path-dependence in games with more than one equilibrium).

[3] A website with a simple CH model "calculator," which calculates CH predictions for ranges of τ for any 2×2 matrix game with up to 50 strategies, is available at http://groups.haas.berkeley.edu/simulations/ch/.

[4] In dynamic games, players are usually assumed to use Bayes's rule to update their beliefs about what will happen next after every observed move. However, when a zero-probability ("out-of-equilibrium") move occurs, Bayes's rule cannot be used. Since all strategies are chosen with positive probability in a QRE, the zero-probability problem never occurs. In a sense, QRE endogenizes the probability of "trembling" used by earlier theorists to resolve the zero-probability problem. In games with mixed-strategy Nash equilibrium, player A's mixture probabilities should only depend on B's payoffs, not on A's. In QRE, both B's payoffs and her own payoffs matter to A. When the response sensitivity parameter $\lambda = 0$, players choose randomly (step-0 thinking). As the response sensitivity parameter λ goes to infinity, choices generally converge to the Nash equilibrium (with some minor exceptions). Given this, perhaps Nash equilibrium could be renamed "hyperresponsive QRE."

4. Learning

Early discussions in game theory were agnostic about how an equilibrium might arise. Recently, theorists have explored the mathematical properties of evolutionary dynamics (e.g., replicator dynamics) and learning rules (Weibull 1995; Fudenberg and Levine 1998). Evolutionary dynamics cannot explain the rapid pace of individual learning in the lab, so I will concentrate on learning rules.

Several rules have been studied. A general rule that fits and predicts well, and includes interesting parametric special cases is called "experience-weighted attraction" (EWA; Camerer and Ho 1999). In EWA, for player i each strategy s_i^j has a level of attraction $A_i^j(t)$, a real number. Attractions are updated in each period to reflect experience according to $A_i^j(t) = [\phi A_i^j(t-1)+(\delta+(1-\delta)I(s_i^j, s_i(t)))\pi_i(s_i^j, s_{-i}(t))]/\phi(1-\kappa)+1)$. (The indicator function $I(x, y)$ is 1 if $x = y$ and zero otherwise; $s_i(t)$ and $s_{-i}(t)$ denote the actual choice by i and other players $(-i)$ in period t.) Attractions map into choice probabilities using a logit response rule.

The term $[\delta+(1-\delta)I(s_i^j, s_i(t))]$ weights the payoff from a strategy. For the strategy s_i^j which was chosen, the indicator function is one so the received payoff gets a weight of one. For strategies which were not chosen, the indicator function is zero so the foregone payoffs of those strategies get a weight of δ.

The EWA theory expresses three features of learning: (1) The decay rate on lagged attractions, ϕ, represents either forgetting or a conscious decision to discard old information when the environment is changing rapidly; (2) δ represents imagination or "regret", the weight on foregone payoffs relative to received payoffs; and (3) κ controls whether attractions average or cumulate, expressing the explore/exploit trade-off in machine learning. A low κ corresponds to continually exploring (because the attractions of strategies are averages and are close together so that the probabilities of choosing them are close, too). A high κ reflects attractions that cumulate, locking in to a good strategy.

EWA hybridizes two approaches that have been widely studied—reinforcement and belief learning. In reinforcement learning, $\delta = 0$ so unchosen strategies are not reinforced (Bush and Mosteller 1983; Arthur 1991; Erev and Roth 1998; Sarin and Vahid 2001). This may reflect ignorance by humans about what they would have received, or cognitive limits (in animal learning). In some forms of reinforcement models, "similar" strategies are reinforced according to the payoffs of chosen strategies. Adjusting responsiveness for how variable reinforcements are also seems to reflect a basic principle of human behavior and improves fit (e.g., Roth et al. 2002).

In belief-learning models, players learn about what others are likely to do, based on their opponents' past choices. For example, in "weighted fictitious play," a player takes an exponentially weighted average of what another player did in the past to guess that player's likely future choice, then uses that belief to calculate expected payoffs from her own strategies (Fudenberg and Levine 1998). Since expected payoffs calculated using this rule are the same as EWA attractions with $\delta = 1$ and $\kappa = 0$, belief learning is simply generalized reinforcement learning in which unchosen strategies are reinforced by foregone payoffs (contrary to implicit claims of behaviorist psychologists for many decades).

Reinforcement and belief models are usually better approximations to the time path of experimental data than equilibrium predictions, for aggregated data (Roth et al. 2000). These simple cases are easy to use because they have few free parameters. However, hybridizing the reinforcement and belief approaches is also a statistical improvement in many of the two dozen or so games that have been studied (although not all). The learning models also add "economic value" in the sense that subjects would have earned more money if they had used the models to forecast what others would do, compared to how much they actually earned (e.g., Ho, Camerer, and Chong 2002).

Figure 13.5 shows predictions of the EWA model fitted to the p-beauty contest data in figure 13.4 (with the initial conditions $A_i^j(0)$ fixed by the data). The model captures the basic tendency of the data to move toward equilibrium. It also improves substantially on the special cases of belief learning and reinforcement. Since most players lose the game and get no reinforcement, simple reinforcement theories predict too little learning. Oppositely, belief theories cannot explain the sluggishness of learning and inertia (captured by $\delta < 1$, so unchosen strategies are not reinforced as strongly as chosen ones).

Current research focuses on whether players learn about learning *rules* (rather than about specific strategies [Salmon 2002]), field applications, and players who are "sophisticated" enough to realize others are learning (Camerer, Ho, and Chong 2002). Sophistication is particularly important if players are matched together repeatedly—like workers in firms, firms in strategic alliances, neighbors, spouses, etc. Then players have an incentive to take actions that teach an adaptive player what to do. Teaching can explain when players behave badly (firms fighting competition to deter further competition) and nicely (to teach others that they can be trusted).

5. Judgment and Choice in Games

Another direction is exploring when systematic deviations from rationality observed in individual choices occur in games as well. For example, it is well known that people expect random series to even out more rapidly then they do; this leads to alternating strategies too often when people play games that require unpredictable randomization (Rapoport and Budescu 1992). There is ample evidence of "framing effects" in which gambles with equivalent dollar payoffs are treated differently when described as gains or losses from different reference points. Extending this possibility to games, when game payoffs are described as losses, players take longer to choose and take more risk (Camerer et al. 1993), are less cooperative (Andreoni 1995), and pass up more mutually beneficial trades (Bazerman 1985), compared to gain-framed games with equivalent final payoffs. Ambiguity-aversion in choices, which corresponds mathematically to a pessimistic reluctance to take action when important information is missing, appears to be present in games: When players'

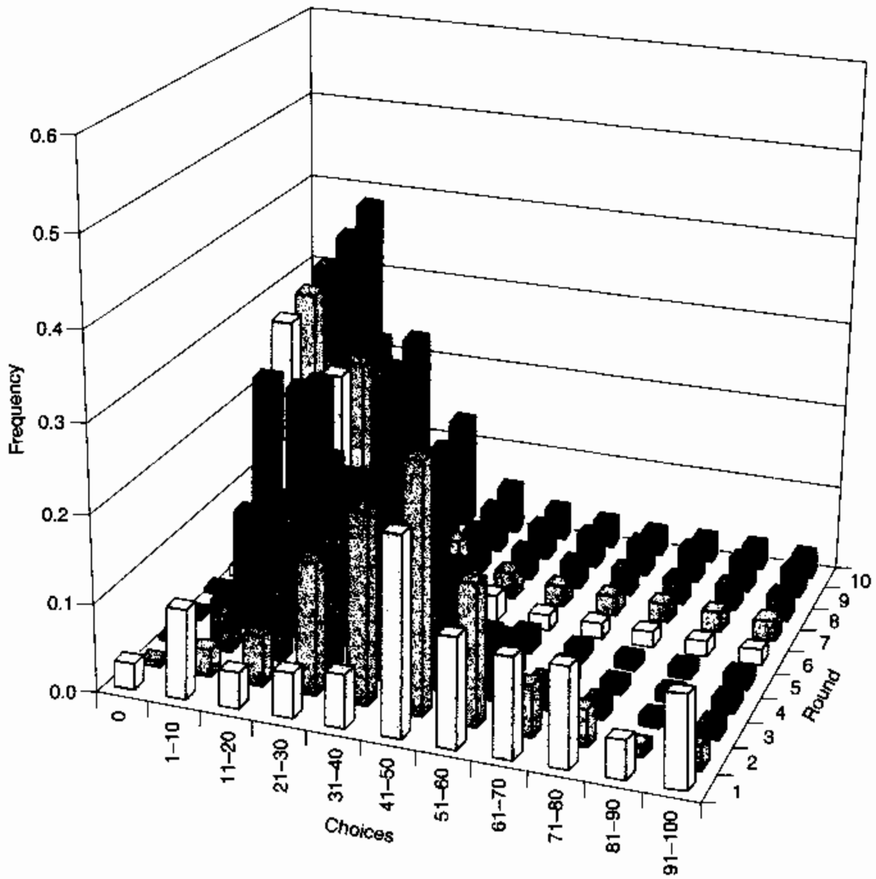

FIGURE 13.5 Fit and prediction of adaptive learning model to number choices in *p*-beauty contest game played 10 times (46).

beliefs about what others will do are measured, they sum to less than one (Camerer and Karjalainen 1994). Many studies show that people (especially men) are overconfident about their relative skill and prospects in life. In competitive games mimicking entry into new businesses, subjects are overconfident (they all think they are more skilled than average, and as a result, lose money as a group) and they neglect the number and skill of likely competitors (Camerer and Lovallo 1999; Moore 2002).

6. Framing, Coordination, and Representation

Framing effects in individual choice are surprising because an invisible axiom of preference theory is that the way a choice is described should not influence its

attractiveness ("description-invariance"). However in games framing can matter—and can *help*—if players desire to coordinate their behavior on one of several norms or equilibria. For example, players in ultimatum games divide less evenly when the game is described as a buyer–seller interaction, or when the Proposer earns the right to make an offer by winning a preplay contest (Hoffman et al. 1994). These description changes appear to evoke different shared social norms for what divisions are fair (à la equity theories in social psychology).

Framing effects are particularly important in games where players have a common interest in coordinating their actions, because the way strategies are described can focus attention on psychologically prominent focal points. Coordination games are an embarrassment for standard theory because it is hard to derive mathematical rules that pick out the one of many equilibria that is obvious (and usually played). Suppose two players can simultaneously choose Red or Blue. They earn $10 if they both choose Red and $5 if they both choose Blue. They will surely choose Red. But both choosing Blue is also a Nash equilibrium. Behavioral theories explain the obvious choice of Red by assuming that players implicitly act as a team (Sugden 2000), or players use a "Stackelberg heuristic": They act as if they are going first, but others will figure out what they are likely to have chosen and "follow" them (Colman and Stirk 1998; Weber and Camerer in press). In game-theoretic jargon, labels and timing are correlating devices that direct shared attention to one of many equilibria; but careful observation is useful for figuring out how these correlating devices work.

A related direction is mental representation. Theorists analyze games in the form of matrices, or trees, but players presumably construct internal representations that may barely resemble matrices or trees. Just as people do not represent explicitly false propositions in mental models of logic, players appear to under-represent payoffs of others in their mental models of games (e.g., Camerer and Johnson, in press; Goldvarg-Steingold and Johnson-Laird 2002). Games with mixed motives, and with conflicting rankings of outcomes across players (biorders), are also difficult to represent (Devetag and Warglien 2002). Limits on representation are particularly important when games are quite complex, with many players and strategies, unfolding over time (like diplomatic maneuvering or planning a business strategy), and are a subject of intense research in multi-agent machine learning.

7. Conclusions

Previous applications of game theory assume that players care only about their own payoffs, and introspect or adapt their way to an equilibrium in which all players mutually best-respond. Experiments show that this simplified model of human behavior in strategic interaction is often violated. The violations point to a general approach, "behavioral game theory," which generalizes standard theory to match observed regularity and psychological intuition. Behavioral game theory combines three ingredients.

The first ingredient is a theory of social utility, which is constructed from evidence about how much players will sacrifice to reduce inequality of payoffs or reciprocate behavior that has helped or hurt them. The second ingredient is a theory of first-period play or initial conditions, which assumes players use different amounts of iterated reasoning, or variants of stochastic "quantal response" equilibria in which players anticipate unpredictable moves by others. The third ingredient is a theory of learning—how experience changes behavior.

Large leaps have been made in the past several years in wrapping the mathematical discipline, which has made game theory so successful in social science applications, around experimental regularity. The next step is to use behavioral theories to make predictions about new games and analyze field phenomena like contract structure, bidding in auctions, industrial competition, social conflict, cooperativeness, bargaining, creations and maintenance of social norms, social capital, and economic growth.

References

Andreoni, James. 1995. "Warm-glow versus cold-prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments." *Quarterly Journal of Economics*, 110: 1–21.

Arthur, Brian. 1991. "Designing economic agents that act like human agents: A behavioral approach to bounded rationality." *American Economic Review*, 81: 353–59.

Aumann, Robert. 1985. "What Is Game Theory Trying to Accomplish?" In *Frontiers of Economics*, edited by K. Arrow and S. Honkaphoja. Oxford: Basil Blackwell.

Bazerman, Max H., Thomas Magliozzi, and Margaret Neale. 1985. "Integrative Bargaining in a Competitive market." *Organizational Behavior and Human Performance* 35: 294–313.

Berg, Joyce, John Dickhaut, and Kevin A. McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, 10: 122–42.

Blount, Sally. 1995. "When Social outcomes Aren't Fair: The Effect of Causal Attributions on Preferences." *Organization Behavior and Human Decision Processes*, 63: 131–44.

Bresnahan, Tim, and Peter Reiss. 1991. "Econometric Models of Discrete Games." *Journal of Econometrics*, 48: 57–81.

Bolton, Gary, and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90: 166–93.

Bush, Robert, and Frederick Mosteller. 1955. *Stochastic Models for Learning*. New York: Wiley.

Camerer, Colin F. 1997. "Progress in Behavioral Game Theory." *Journal of Economic Perspectives*, 11: 167.

Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton: Princeton University Press.

Camerer, Colin, and Teck Ho. 1999. "Experience-Weighted Attraction Learning in Normal-Form Games." *Econometrica*, 67: 827–74.

Camerer, Colin, Teck Ho, and Kuan Chong. 2002. "Sophisticated EWA Learning and Strategic Teaching in Repeated Games." *Journal of Economic Theory*, 104: 137–88.

———. 2003. "Models of Thinking, Learning and Teaching in games." *American Economic Review*, 93(2) (May): 192–95.

Camerer, Colin F., and Eric J. Johnson. In press. "Thinking about Attention in Games." In *The Psychology of Economic Decisions*, Vol. 2, edited by I. Brocas and J. Castillo. Oxford: Oxford University Press.

Camerer, Colin F., Eric Johnson, Talia Rymon, and Sankar Sen. 1993. "Cognition and Framing in Sequential Bargaining for Gains and Losses." In *Frontiers of Game Theory*, edited by Ken Binmore, Alan Kirman, and Piero Tani. Cambridge: MIT Press.

Camerer, Colin F., and Risto Karjalainen. 1994. "Ambiguity-Aversion and Non-additive Beliefs in Non-cooperative Games: Experimental Evidence." In *Models and Experiments on Risk and Rationality*, edited by Bertrand Munier, and Mark Machina. Dordrecht: Kluwer.

Camerer, Colin, and Dan Lovallo. 1999. "Overconfidence and Excess Entry: Experimental Evidence." *American Economic Review*, 89: 306–18.

Camerer, Colin, and Richard Thaler. 1995. "Anomalies: Dictators, Ultimatums, and Manners." *Journal of Economic Perspectives*, 9: 209–19.

Colman, Andrew. In press. "Cooperation, Psychological Game Theory, and Limitations of Rationality in Social Interaction." *Behavioral and Brain Sciences*.

Colman, Andrew, and John A. Stirk. 1998. "Stackelberg Reasoning in Mixed-Motive Games: An Experimental Investigation." *Journal of Economic Psychology*, 19: 279–93.

Cooper, Russell. 1999. *Coordination Games: Complementarities and Macroeconomics*. Cambridge: Cambridge University Press.

Costa-Gomes, Miguel, Vincent Crawford, and Bruno Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica*, 69: 1193–235.

Crawford, Vincent. 1997. "Theory and Experiment in the Analysis of Strategic Interaction." In *Advances in Economics and Econometrics: Theory and Applications*, edited by D. Kreps and K. Wallis. Cambridge: Cambridge University Press.

Cross, John. 1983. *A Theory of Adaptive Economic Behavior*. New York and London: Cambridge University Press.

Dana, Jason, Roberto Weber, and Jason Kuang. 2003. "Being Fair vs. Not Being a Jerk: Fragile Preferences for Social Outcomes." Working paper, Carnegie-Mellon University.

Devetag, Giovanna, and Massimo Warglien. 2002. "Representing others' Preferences in Mixed Motive Games: Was Schelling Right?" Unpublished paper, University of Trento.

Erev, Ido, and Alvin Roth. 1998. "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed-Strategy Equilibria." *American Economic Review*, 88: 848–81.

Fehr, Ernst, and Klaus Schmidt. 1999. "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*, 114: 817–68.

Fudenberg, Drew, and David Levine. 1998. *The Theory of Learning in Game*. Cambridge: MIT Press.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge: MIT Press.

Gale, John, Kenneth Binmore, and Larry Samuelson. 1995. "Learning to Be Imperfect: The Ultimatum Game." *Games and Economic Behavior*, 8: 59–60.

Goeree, Jacob, and Charles Holt. 1999. "Stochastic Game Theory: For Playing Games, Not Just for Doing Theory." *Proceedings of the National Academy of Sciences*, 96: 10564–567.

———. 2001. "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions." *American Economic Review*, 91: 1402–422.

Goldvarg-Steingold, Eugenia, and Philip N. Johnson-Laird. 2002. "Naïve Strategic Thinking and the Representation of Games." Working paper, Princeton University.

Gunnthorsdottir, Anna, Kevin McCabe, and Vernon Smith. 2002. "Using the Machiavellian Instrument to Predict Trustworthiness in a Bargaining Game." *Journal of Economic Psychology*, 23: 49–66.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*, 3: 367–88.

Harsanyi, John. 1967a. "Games with Incomplete Information Played by Bayesian Players I." *Management Science*, 14: 159–82.

———. 1967b. "Games with Incomplete Information Played by Bayesian Players II." *Management Science*, 14: 320–34.

———. 1968. "Games with Incomplete Information Played by Bayesian Players III." *Management Science*, 14: 486–502.

Hendricks, Ken, and Harry Paarsch. 1995. "A Survey of Recent Empirical Work Concerning Auctions." *Canadian Journal of Economics*, 28: 403.

Henrich, Joe. 2000. "Does Culture Matter in Economic Behavior? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon." *American Economic Review*, 90: 973–79.

Henrich, Joe, Robert Boyd, Sam Bowles, Herb Gintis, Ernst Fehr, Colin F. Camerer, and Richard McElreath. 2000. "Cooperation, Reciprocity and Punishment: Experiments in 15 Small-scale Societies." Discussion paper, University of Michigan.

Ho, Teck, Colin Camerer, and Kuan Chong. 2002. "Functional EWA: A One-Parameter Theory of Learning in Games." Working paper, Caltech.

Ho, Teck, and Kuan Chong. 2003. "A Parsimonious Model of SKU Choice: Familiarity based Reinforcement and Response Sensitivity." *Journal of Marketing Research*, 40(3), (forthcoming).

Ho, Teck, Colin Camerar, and Keith Weigelt. 1998. "Iterated Dominence and Iterated Best-response in *p*-Beauty Contests." *American Economic Review*, 88: 947–69.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. 1994. "Preferences, Property Rights and Anonymity in Bargaining Games." *Games and Economic Behavior*, 7: 346–80.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith. 1996. "On Expectations and Monetary Stakes in Ultimatum Games." *International Journal of Game Theory*, 25: 289–301.

Kennan, John, and Robert Wilson. 1990. "Theories of Bargaining Delays." *Science*, 249: 1124–28.

Keynes, John Maynard. 1936. *The General Theory of Interest, Employment, and Money*. London: Macmillan.

Kirman, Alan, and Nicolaas Vriend. 2000. "Learning to Be Loyal: A Study of the Marseille Fish Market." In *Interaction and Market Structure. Essays on Heterogeneity in Economics*, edited by D. Delli Gatti, M. Gallegati, and A. P. Kirman. Berlin: Springer.

Laffont, Jean-Jacques. 1997. "Game Theory and Empirical Economics: The Case of Auctions." *European Economic Review*, 41: 1–35.

Maynard Smith, John. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

McAfee, Preston, and John McMillan. 1996. "Analyzing the Airwaves Auction." *Journal of Economics Perspectives*, 10: 159–75.

McKelvey Richard, and Thomas Palfrey. 1992. "An Experimental Study of the Centipede Game." *Econometrica*, 60: 803–36.

———. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior*, 7: 6–38.

———. 1998. "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics*, 1: 9–41.

Milgrom, Paul Milgrom, and John Roberts. 1992. *Economics, Organization and Management*. Englewood Cliffs, N.J.: Prentice-Hall.

Moore Don. 2002. "Strategic Foresight in Market Entry Decisions: An Experimental Approach." Working paper, Carnegie-Mellon University.

Nash, John. 1950. "Equilibrium Points in N-person Games." *Proceedings of the National Academy of Sciences*, 36: 48.

Nagel, Rosemarie. 1995. "Unraveling in Guessing Games: An Experimental Study." *American Economic Review*, 85: 1313–26.

———. 1999. "A Review of Beauty Contest Games." In *Games and Human Behavior: Essays in honor of Amnon Rapoport*, edited by D. Budescu, I. Erev, and R. Zwick. Hillsdale, N.J.: Lawrence Erlbaum.

Nowak, Martin, Karen Page, and Karl Sigmund. 2000. "Fairness versus Reason in the Ultimatum Game." *Science*, 289: 1773–75.

Osborne, Martin, and Ariel Rubinstein. 1995. *A Course in Game Theory*. Cambridge: MIT Press.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature*, 37: 7–63.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83: 1281–302.

Rapoport, Amnon, and David V. Budescu. 1992. "Generation of Random Series in Two-Person Strictly Competitive Games." *Journal of Experimental Psychology: General*, 121: 352–63.

Roth, Alvin, Ido Erev, Robert Slonim, and Greg Barron. 2002. "Equilibrium and Learning in Economic Environments: The Predictive Value of Approximations." Discussion paper, Department of Economics, Harvard University.

Roth, Alvin, and Elliott Peranson. 1999. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design." *American Economic Review*, 89: 748–80.

Roth, Alvin, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir. 1991. "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An Experimental Study." *American Economic Review*, 81: 1068–95.

Salmon, Tim. 2001. "An Evaluation of Econometric Models of Adaptive Learning." *Econometrica*, 69: 1597–628.

Samuelson, Larry. 2000. "Analogies, Adaptation, and Anomalies." *Journal of Economic Theory*, 97: 320–66.

Sarin, Rakesh, and Farsid Vahid. 2001. "Predicting How People Play Games: A Simple Dynamic Model of Choice." *Games and Economic Behavior*, 34: 104–22.

Selten, Reinhard. 1975. "Re-examination of the perfectness concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory*, 4: 25.

Shepsle, Kenneth, and Mark Boncheck. 1997. *Analyzing Politics*. New York: W.W. Norton.

Stahl, Dale. 1993. "Evolution of *Smart$_n$* Players." *Games and Economic Behavior*, 5: 604–17.

———. 2000a. "Rule Learning in Symmetric Normal-Form Games: Theory and Evidence." *Games and Economic Behavior*, 32: 105–38.

———. 2000b. "Sophisticated Learning and Learning Sophistication." Discussion paper, Department of Economics, University of Texas.

Stahl, Dale, and Paul Wilson. 1995. "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, 10: 218–54.
Sugden, Robert. 2000. "Team preferences." *Economics and Philosophy*, 16: 175–204.
Tirole, Jean. 1998. *The Theory of Industrial Organization*. Cambridge: MIT Press.
Von Neumann, John and Oskar Morgenstern. 1944. *The Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
Weber, Roberto, Colin F. Camerer, and Marc Knez. In press. "First-Mover Advantage and Virtual Observability in Ultimatum Bargaining and Weak-link Coordination Games." *Experimental Economics*.
Weibull, Jörgen. 1995. *Evolutionary Game Theory*. Cambridge: MIT Press.