Journal of Economic Theory **104**, 1–15 (2002)
doi:10.1006/jeth.2001.2909, available online at http://www.idealibrary.com on IDEAL®

Introduction to Experimental Game Theory

Vincent P. Crawford[1]

Department of Economics, University of California at San Diego, La Jolla, California 92093-0508
vcrawfor@weber.ucsd.edu

Journal of Economic Literature Classification Numbers: C70, C92.
Key Words: game theory; experimental economics; behavioral economics; strategic behavior.

1. INTRODUCTION

The title Experimental Game Theory refers to experiments whose goal is to learn about general principles of strategic behavior, as opposed to the performance of specific institutions. Twenty-five years ago it would have been startling to see "experimental" modifying "game theory" in this way or to find an entire issue of an economics journal devoted to game experiments.[2] If this now seems natural, it is a tribute to the increasingly empirical orientation of game theory and the researchers who have made experiments an important tool for the analysis of strategic behavior.[3]

With hindsight, the emergence of experimental game theory can be traced to two factors—the need for empirical information about principles of strategic behavior and the advantages of experiments in providing it.

[1] I am grateful to the National Science Foundation for research support and to Colin Camerer, Miguel Costa-Gomes, and Larry Samuelson for helpful advice.

[2] See, however, the 1995 *Games and Economic Behavior* special issue on experimental game theory introduced by Ledyard and Palfrey [26]. Twenty-five years ago, the only experiment that had commanded the attention of the entire profession was Smith's [41] classic study of the competitive double-oral-auction market. His experiment revealed—at a time when most economists thought perfect competition required many traders, with perfect information—that markets with only a few traders on each side can yield competitive outcomes and that market outcomes may actually be more competitive when traders are imperfectly informed. The widespread impact of Smith's results was probably due largely to their robustness to the details of individual behavior—details that are the focus of experimental game theory.

[3] Kagel and Roth [23], Crawford [12, 13], Selten [39], and Camerer [3, 4] provide complementary surveys of the experimental literature, and Plott [30], Smith [42, 43], and Roth [34, 35] discuss its methodology.

0022-0531/02 $35.00

This need was long obscured by the game-theoretic custom of trying to predict behavior entirely by theory, applying notions of equilibrium and refinements to the structure of the game.[4] This custom admits a role for empirical knowledge about players' preferences, feasible decisions, and information just as in nonstrategic microeconomics, but it precludes any role for empirical input about the principles that determine how players respond to a given game.

Excluding such input is comparatively innocuous in nonstrategic settings, where rationality in the sense of expected-utility maximization often provides a reasonably reliable guide to behavior once preferences, decisions, and information have been identified. But it is far from innocuous in strategic settings, where rationality alone seldom yields definite predictions, reliable or not, and consensus about how to strengthen it is if anything more remote than 25 years ago. As a result, most games of interest in economic applications raise questions about strategic behavior that seem likely to be adequately resolved only by combining theory and empirical knowledge.[5]

The need for empirical knowledge about principles of strategic behavior creates a special role for experiments in game theory. The predictions of game theory—particularly noncooperative game theory, which underlies most applications—are notoriously sensitive to the details of the structure of the game, and much of this sensitivity is reflected in observed behavior. Such details can seldom be precisely observed or adequately controlled in the field. The laboratory shares some of these problems, but the control and observation that modern experimental techniques allow often give experiments a decisive advantage in identifying the relationship between strategic behavior and the environment.[6] In this endeavor, theory and experiment play strongly complementary roles, with theory providing a framework within which to gather and interpret empirical information about behavior, and experiments indicating which parts of the theory are

[4] The custom may reflect the early conception of game theory as a mathematical investigation of the behavior of idealized "perfectly rational" agents and the view that this would be fully adequate to predict strategic behavior. Schelling [37] gave an early and influential dissenting view, which did little to alter the custom among theorists.

[5] Economic games usually have multiple rationalizable outcomes and often have multiple equilibria. In principle, the limitations of a purely rationality-based approach could be overcome by more powerful theory, as for example in Harsanyi and Selten [20]; but this program now appears unlikely to succeed without admitting empirical knowledge. The limitations are transformed, but not eliminated, by viewing equilibrium as the outcome of an adaptive learning process, which raises many new issues that theory alone does not satisfactorily resolve.

[6] There is nonetheless a history of valuable empirical work using field data from strategic environments, usually with well-specified, observable structures, as for example in auctions or centralized labor markets.

most useful in predicting behavior, and identifying behavioral parameters that theory does not reliably determine.

This symposium showcases some of the best recent work in experimental game theory, work that highlights the power of experimental methods to elucidate questions central to game theory and illustrates the possibilities of experimental design. The rest of this introduction describes the papers, which are broadly grouped by game-theoretic topic.[7] Section 2 introduces the first four papers, which present primarily static analyses of behavior in extensive-form games: backward induction, social preferences, implementation, and preplay communication. Section 3 introduces the next three papers, which present dynamic analyses of learning in normal-form games: strategic teaching, learning mixed strategies, and analogies. Section 4 introduces the final paper, which considers the possibility of using a quantal response equilibrium model with risk aversion to explain overbidding (relative to the risk-neutral Nash equilibrium) in first-price sealed-bid auctions.

2. BACKWARD INDUCTION, SOCIAL PREFERENCES, IMPLEMENTATION, AND PREPLAY COMMUNICATION IN EXTENSIVE-FORM GAMES

Johnson, Camerer, Sen, and Rymon [22] report experiments that elicit subjects' initial responses to a series of two-person three-period alternating-offers bargaining games, in which the effects of discounting are simulated by shrinking the size of the "pie" to be divided over time, with different pie sizes and partners each period to suppress learning and repeated-game effects.[8] If players maximize pecuniary payoffs and information is complete, these games have unique subgame-perfect equilibria that are easily computed by backward induction. Previous experiments have yielded large, systematic deviations from the subgame-perfect equilibrium offer and acceptance decisions, like those usually observed in ultimatum experiments. These deviations have been attributed to cognitive limitations that prevent subjects from doing the required backward induction (or from believing their partners will, etc.) or to subjects having "social" preferences that

[7] I make no systematic attempt to explain the methods of experimental game theory, instead referring interested readers to the surveys and discussions mentioned above and the expositions in the papers themselves.

[8] Designs that elicit initial responses are comparatively unusual in experimental game theory, where having subjects play the same game repeatedly often reduces the noisiness of their responses as they learn the game from experience. Eliciting initial responses can help to identify strategic principles because repeated play of the same game often converges to equilibrium no matter what subjects are thinking. However, by foregoing repetition as a teaching device, such designs place a heavier burden on subjects' understanding, with a premium on simplicity and clarity of design.

respond to the perceived fairness of outcomes as well as pecuniary payoffs.[9] Most researchers now agree that both factors are significant, but their relative importance has been hard to determine.

This question, like many that concern strategic behavior, turns partly on cognition. In economics cognition is usually studied only indirectly, by inference from the model that best describes observed decisions. With careful design this can make it possible to infer what subjects must have been thinking, but it would plainly be useful to study cognition more directly. Johnson *et al.* do this by monitoring subjects' searches for hidden but freely available information about payoffs, using experimental software originally developed for research on individual decisions.

Their design has three treatments, in which the same games are presented to (mostly) different subjects. In the baseline treatment, subjects played the games against each other and were rewarded according to their game payoffs. In a "robot" control treatment, subjects played the games against the computer, which was programmed to follow its subgame-perfect equilibrium strategy. These subjects also received training in backward induction after the first few periods. In a third treatment, subjects who had been through the robot treatment, with training, were paired with untrained subjects, and all were again rewarded according to their game payoffs. In each case, everything about the environment was publicly announced but the pie sizes, to which subjects were given unlimited access through a MouseLab computer interface, which automatically recorded their information searches along with their decisions.[10] The design thus allowed subjects to evaluate their own and their partners' pecuniary payoffs for any decision combination, making the structure of the games effectively *public knowledge*, except possibly for subjects' preferences.

Johnson *et al.* argue that backward induction has a characteristic search pattern, in which subjects first look up the last-period pie size, then the second-last, and so on, with most transitions from later to earlier periods. They supported this claim empirically by showing that subjects trained in backward induction tend to exhibit such a pattern.[11] By contrast, untrained baseline subjects deviated systematically from the backward-induction search pattern, and subjects whose patterns were closer to backward

[9] Even with privately observed social preferences, backward induction yields a generically unique equilibrium.

[10] For more information about MouseLab, see http://ecom.gsb.columbia.edu/mouselab/MouseLab.htm. In Johnson *et al.*'s design, varying the games serves the additional purpose of making it impossible for subjects to remember the pie sizes from previous plays, which would weaken the link between cognition and information search.

[11] See Camerer *et al.* [7], which reports experiments with a structurally equivalent set of games, whose primary goal was comparing an "expanding-loss" framing with the "shrinking-gains" framing studied here.

induction tended to make offer and acceptance decisions closer to the subgame-perfect equilibrium. Johnson *et al.* also argue that the "robot" treatment disables social preferences, so that their importance can be estimated by comparing its results with those of the baseline. Their results allow a more precise assessment of the relative importance of limited cognition and social preferences, which suggests that limited cognition is a more significant factor than one might have thought, based on previous analyses of decisions alone. Their analysis also establishes monitoring information search as a useful, tractable tool for learning about cognition in games, which has many potential applications in studying other aspects of strategic behavior.[12]

Binmore, McCarthy, Ponti, Samuelson, and Shaked [2] report experiments with two-person alternating-offers bargaining games, which like Johnson *et al.*'s experiments are designed to elucidate why subjects fail to play the subgame-perfect equilibrium in which players maximize pecuniary payoffs. Binmore *et al.* start with a careful discussion of possible theoretical explanations, dissecting backward induction into its component assumptions of rationality, subgame consistency (play in a subgame is independent of its position in a larger game), and truncation consistency (replacing a subgame with its equilibrium payoffs does not affect play elsewhere in the game). They then use the theory to create a design that allows them to identify the sources of failures to play the subgame-perfect equilibrium by direct comparisons across games. In their treatments, each subject plays a series of 20 plays each of four different two-person one- and two-period alternating-offers bargaining games, with different partners each period to suppress repeated-game effects. The games are chosen to isolate failures of subgame consistency and truncation consistency individually or in combination, assuming rationality (but not pecuniary payoff maximization) and controlling for social preferences that depend only on subjects' own and other subjects' monetary payoffs.[13]

Binmore *et al.*'s results for individual games replicate those of previous experiments. Their comprehensive design, together with a detailed econometric analysis that takes advantage of the panel structure of the data,

[12] In work directly inspired by Camerer *et al.*'s and Johnson *et al.*'s analyses, Costa-Gomes *et al.* [11] study two-person matrix games, using MouseLab to allow subjects to look up their own and their partners' payoffs for each decision combination. In this design subjects have a larger space of possible search patterns, and their deviations from the search patterns suggested by equilibrium analysis help to explain their deviations from equilibrium decisions.

[13] Binmore *et al.* call such preferences "payoff-interdependent." The leading theories of payoff-interdependent preferences are consistent with rationality by construction. Binmore *et al.* rule out more general kinds of social preferences (e.g, payoffs that depend on the strategies that led to them, the structure of the game, etc.) because without some restriction such as theirs, backward induction has few refutable implications.

makes it possible to identify the most likely sources of subjects' deviations from the subgame-perfect equilibrium in which players maximize pecuniary payoffs. They find evidence of social preferences, as in previous experiments with such games; but they also find widespread, systematic violations of subgame consistency ("proposers are less aggressive in the second stage of a two-stage bargaining game than in an equivalent one-stage game") and truncation consistency ("[p]layers are less responsive to variations in the expected value of playing a subgame than to equivalent variations in terminal payoffs"), undetected in previous experiments that did not separate the assumptions that underlie backward induction.

Katok, Sefton, and Yavas [25] report experiments that advance our understanding of the role of extensive-form games in the implementation of social choice functions. Their work was motivated by the Abreu–Matsushima [1] mechanism, a simultaneous-move mechanism that has played a leading role in the implementation literature, which virtually implements a wide range of social choice functions in iteratively undominated strategies.[14] For this mechanism to perform as intended, players must respect a large number of rounds of iterated strict dominance. Experimental evidence from many other kinds of games suggests that this is unlikely without extensive prior experience with the specific game in question: In their initial responses to games, subjects seldom play dominated strategies, but usually respect at most three or four rounds of iterated dominance, far too few for the Abreu–Matsushima mechanism to perform as predicted in theory (see for example, Glazer and Rosenthal [17]). Sefton and Yavas [38] report experiments that confirm this for the game induced by the Abreu–Matsushima mechanism, using a design based on the simple implementation problem Glazer and Rosenthal used to illustrate the mechanism.

Glazer and Perry [16] have recently proposed a sequential version of the Abreu–Matsushima mechanism, which replaces its complex iterated dominance argument with a series of simple decisions, with the goal of making the desired equilibrium more transparent.[15] Katok *et al.* [25] replicate Sefton and Yavas' results for the simultaneous version of the mechanism, using a slightly different design, and then compare the performance of the simultaneous and sequential versions. They find, somewhat surprisingly, that the sequential version performs if anything slightly worse than the simultaneous version, although the difference is not statistically significant. Their paper is noteworthy for the clarity of their design and the skill with

[14] "Virtual implementation" here means that there is a social choice function arbitrarily close to the one in question that is exactly implementable in iteratively undominated strategies

[15] In effect, the sequential version tells players the most efficient order in which to compare their strategies.

which they present the complex games involved to their subjects, ensuring their comprehension.

Costa-Gomes [10] reconsiders the results of Cooper *et al.*'s [9] experiments on structured preplay communication in the Battle of the Sexes game and Roth, Malouf, and Murnighan's unstructured bargaining experiments (summarized in Roth [33]), in the light of Rabin's [31] analysis of preplay communication. In Rabin's model communication is structured, with players repeatedly and simultaneously sending non-binding messages about their intended strategies, in a commonly understood language, before playing a game. Using non-equilibrium notions in the spirit of rationalizability and additional, behaviorally motivated restrictions on players' strategies, Rabin shows that in the limit as communication becomes "abundant" (that is, the number of rounds grows without limit), each player is assured an expected payoff at least as great as in his worst Pareto-efficient Nash equilibrium in the underlying game.

Costa-Gomes begins by discussing the correspondence between Rabin's model and the experimental designs. In Cooper *et al.*'s experiments communication was structured, with subjects limited to zero, one, or three rounds of simultaneous messages, but the environment is otherwise quite close to Rabin's model. In Roth *et al.*'s experiments communication was unstructured, with subjects allowed to send essentially unrestricted messages to each other at any time (with a small delivery lag) for a fixed period of time, either 10 or 12 minutes, and the only rules were that any feasible agreement reached before the time limit would be enforced. Two problems are apparent in applying Rabin's analysis: It restricts only limiting outcomes, but in the experiments the communication possibilities are necessarily bounded; and Roth *et al.*'s unstructured communication is superficially quite different from the structured communication in Rabin's model. Costa-Gomes addresses the latter problem by arguing (following Schelling [37, Appendix B] and Harsanyi and Selten [20, pp. 23–24]) that in this setting, unstructured but bounded communication is in fact naturally modeled by structured, simultaneous communication. He addresses the former problem by adding plausible restrictions on players' strategies (still without assuming equilibrium) that allow the model to predict the qualitative effects on players' expected payoffs of changing the number of rounds of structured communication or the time limit in unstructured bargaining. This strengthened version of Rabin's model yields predictions consistent with both sets of experimental results; and under plausible assumptions, Rabin's expected payoff bounds are usually satisfied even though the communication possibilities were bounded. The analysis identifies a link between two seemingly unrelated sets of experimental results and a theory proposed independently, increasing the informativeness of all three contributions.

3. LEARNING, STRATEGIC TEACHING, MIXED STRATEGIES, AND ANALOGIES IN NORMAL-FORM GAMES

The next three papers consider different aspects of learning in normal-form games: strategic teaching, learning mixed strategies, and learning from imperfect analogies.

Camerer, Ho, and Chong [6] propose a simple model to explain a phenomenon they call "strategic teaching," in which a player seeks to benefit by deviating from his short-run payoff-maximizing strategy to influence other players' future decisions. Such benefits are negligible in many experimental settings, for example with repeated random pairing from a "large" population to play two-person games, where subjects quite sensibly ignore them. But some experimental results seem inexplicable without considering something like strategic teaching.

Consider, for instance, Van Huyck *et al.*'s [45] experimental treatments involving two-person minimum-effort coordination games.[16] In those games subjects repeatedly and simultaneously chose among seven efforts, with payoffs determined by their own effort and their pair's minimum effort. The games have seven symmetric Pareto-ranked pure-strategy equilibria, with all players preferring the one in which both choose the highest effort. Subjects were told their pair's minimum after each play, but nothing else about other subjects' efforts. The two treatments were identical, except that in treatment C_d (d for "different") the subjects were randomly re-paired (from a population of either 14 or 16) for each period in a run of either three or five periods, whereas in treatment C_f (f for "fixed") their pairings (from a population of either 12 or 16), while random, were fixed for an entire run of seven periods. In each case, the structure of the environment was publicly announced, including whether or not the pairings were fixed and the fact that the subject population was fixed for the duration.

This difference in pairing schemes led to very different outcomes (Van Huyck *et al.* [45, Tables 4–5]). In treatment C_d, subjects' efforts were widely dispersed, with mean between 4 and 5, moderately inefficient outcomes, and no apparent convergence or time trend. In treatment C_f, by contrast, within seven periods 12 of 14 pairs increased the minimum effort to its fully efficient level, often starting from much lower levels. Treatment C_f's subjects were evidently well aware that they could influence their partners' future efforts, and many exploited this influence; but treatment C_d's subjects apparently treated such influences as negligible.

[16] Those treatments are only a small part of Van Huyck *et al.*'s experiments. Although Camerer *et al.* mention them in passing, their main focus is on analyzing the results of different sets of experiments, as explained below.

What are subjects doing in treatment C_f, and why do they not do it in treatment C_d? It is difficult to give an equilibrium answer to this question. One can easily construct repeated-game equilibria in which C_f players bring about efficient coordination (or achieve it from the start) and C_d subjects do not, but equilibrium in the repeated game is consistent with any time pattern of pair minima, and it is just as easy to reverse these conclusions.[17] Thus, while equilibrium in the repeated game is roughly consistent with the results, it cannot be said to explain them. It is also difficult to give an adaptive learning answer. If players focus on stage-game strategies, the only difference between the treatments is that a C_f subject observes the effort of another subject who will be his partner in the future, while a C_d subject observes the effort of one with whom he will never interact again. This might affect the speed of convergence, but it cannot explain the large difference in likely limiting outcomes that was observed. If players focus instead on strategies with memory, the explanation would ultimately have to rest on ad hoc memory restrictions, without which players could not adjust their strategies within a single play of the repeated game.

Camerer *et al.* cut through these difficulties by allowing a small degree of heterogeneity in players' behavior rules. Some players, they assume, are myopic adaptive learners, who follow Camerer and Ho's [5] experience-weighted attraction (EWA) learning rule. Others are sophisticated and forward-looking, choosing stage-game strategies that best respond to the population, on the assumption that it is composed of a mixture of adaptive and sophisticated subjects. In effect, the sophisticated players play an equilibrium in a game among themselves, with the adaptive learners (whose behavior is a mechanical response to history) treated as part of the game. In this model, sophisticated C_f subjects have a strong incentive to forego current gains to "teach" their possibly adaptive partners to increase effort, but sophisticated C_d subjects have no such incentive.[18]

[17] The treatments can both be viewed as games played by the entire population, with the expected payoffs of players' strategies evaluated before the uncertainty of pairing is resolved. From this point of view, even treatment C_d may allow non-trivial repeated-game equilibria via "contagious" influences as in Kandori [24], although some experiments use pairing schemes that eliminate this possibility.

[18] A sophisticated C_f subject's incentive is weakened by the fact that his partner does not observe the subject's effort, only the pair minimum and his own effort, so that all a sophisticated subject could do was to ensure that his partner's effort equaled the pair minimum. That even this information was enough makes the phenomenon more striking. However, this application involves a difficulty that Camerer *et al.*'s applications avoid. If the proportion of sophisticated players is high enough, their benefits from coordinating with each other in the stage game will outweigh the benefits of teaching adaptive learners, and the game played among sophisticated players will inherit the multiplicity of equilibria of the original coordination game. In this case something else is needed to close the model.

This kind of model, originally mentioned as a theoretical possibility by Fudenberg and Levine [15, pp. 261–263], takes on new life when Camerer *et al.* use it to reconsider the data from Ho *et al*'s [21] experiments with *p*-beauty contest games and Camerer and Weigelt's [8] experiments with repeated borrower–lender trust games. The model determines behavior up to parameters that represent the fractions of naïve learners and sophisticated players, sophisticated players' estimate of those fractions, and the weight sophisticated players assign to the future, which are estimated in each application. In the borrower–lender trust games, the estimated model can be viewed as a way of endogenizing the behavior of the "crazy" type in an equilibrium reputation model like those that originally motivated Camerer and Weigelt's experiment. Camerer *et al.* distinguish the models' observable implications and compare them econometrically, finding that the strategic teaching model outperforms the equilibrium reputation model in the trust games.[19]

Shachat [40] reports experiments that reexamine whether subjects can learn to play unique mixed-strategy equilibria in zero-sum two-person matrix games, a perennial theme in the experimental game theory literature. Here, theory makes a simple, clear prediction that is beyond the intuition of most people not trained in game theory. This makes it natural to ask whether learning will lead them to equilibrium, and this question is readily amenable to experimental study.

The modern literature on this question begins with O'Neill [29], and in the 15 years since his paper there have been a number of further experimental studies. These studies typically reveal systematic deviations from equilibrium: Most subjects' chosen pure strategies are both positively serially correlated and contemporaneously correlated across subject pairs, and the variance of "win rates" (rates of high-payoff outcomes) across subject pairs is higher than the theory predicts. Despite these deviations (or possibly because of them), population aggregate strategy frequencies are invariably much closer to the theory's predictions than individual subjects' strategy frequencies.

These results raise several questions, including whether subjects' strategy choices are serially correlated because they cannot generate independently and identically distributed sequences or because they do not use equilibrium mixed strategies, whether the excess variance of win rates is due solely to the contemporaneous correlation of actions across subject pairs, and why the theory works better in the aggregate than at the individual level. Moreover, despite the obvious advantages of allowing subjects explicitly to choose mixed strategies in this context, previous studies have not done so (with exceptions whose designs do not correspond precisely to the theory).

[19] In estimating the equilibrium reputation model, they allow for the noisiness of subjects' decisions using McKelvey and Palfrey's [28] extensive-form quantal response equilibrium.

Shachat begins with a comprehensive discussion of previous findings, problems with previous designs, and open questions. He then constructs a design that avoids the problems and addresses the questions and that features new software (the "mixed strategy device") that allows subjects to choose mixed strategies almost as easily and transparently as they can choose pure strategies. In his design, as in most previous studies, a subject repeatedly played the same game against the same randomly selected opponent, in this case for 60 periods.[20] As in most previous work, the game has only two possible payoff outcomes, which under plausible assumptions makes the theory's predictions independent of subjects' risk preferences.

After replicating previous results, Shachat finds that the mixed strategy device reduces but does not eliminate the serial correlation in individual subjects' strategy choices and that the serial correlation arises because subjects condition their play on their partners' past play, contrary to equilibrium predictions. He also finds that non-face-to-face interaction reduces but does not eliminate the contemporaneous correlation in subject pairs' action choices and that subjects in the population use a range of pure and/or mixed strategies, which are typically not equilibrium strategies and whose deviations from equilibrium add to the excess variance of win rates.

Van Huyck and Battalio [44] study learning from imperfect analogies in a class of discrete bargaining games. To date, almost all analyses of learning, theoretical or experimental, have concerned learning to play a single, fixed game, with past plays perfectly analogous to present ones, and past behavior taken to be directly representative of likely present behavior.[21] Real analogies are seldom this perfect, and how players learn from others' behavior in games that are similar but not identical is an important open question. Such learning requires players to interpret their experience using general principles that create analogies between games. This, of course, is just what game-theoretic solution concepts do; but existing theory seems unlikely to provide an adequate description, because no concept yet proposed has commanded wide acceptance, and the leading ones (e.g. Harsanyi and Selten [20]) seem too complex to describe human behavior.

Rankin *et al.* [32] opened an investigation of this issue by studying subjects' limiting behavior in repeated play of similar Stag Hunt games. They disabled mechanical learning by perturbing the games' payoffs and action labels each period, so that subjects could use their experience in previous plays only via deductive analogies between games. The results were surprisingly clear: Seven of seven subject groups either converged or

[20] Because the games are zero-sum with unique equilibria, the only equilibrium (subgame-perfect or not) in the repeated game that describes a pair's interaction is repeated play of the equilibrium in the stage game, independent of history.

[21] Samuelson [36], in which analogies emerge endogenously in an evolutionary setting, is a recent theoretical exception.

appeared to be converging (within 56 or 75 periods) to the payoff-dominant equilibrium, in contrast to previous results for repeated play of identical Stag Hunt games, which strongly favored risk-dominance.

Van Huyck and Battalio continue this investigation in 2×2 bargaining games with two Pareto-efficient strict equilibria, one favored by the utilitarian criterion and one favored by Rawls' maximin criterion, using techniques like Rankin *et al.*'s to disable mechanical learning. They find emergent conventions (within 70 periods) based on deductive principles in 5 of 26 subject groups, 4 utilitarian and 1 Rawlsian.

The nature of the analogies between games that inform Rankin *et al.*'s and Van Huyck and Battalio's subjects' behavior is still largely an open question, and their results are an intriguing challenge for theorists and experimentalists.

4. QUANTAL RESPONSE EQUILIBRIUM IN FIRST-PRICE SEALED-BID AUCTIONS

Goeree, Holt, and Palfrey [18] report experiments that revisit a controversial issue in the experimental auction literature: whether risk aversion can explain the common tendency of human subjects in first-price sealed-bid auctions with private values to "overbid" relative to the risk-neutral Nash equilibrium. In response to Harrison's [19] "flat-maximum" critique of the original overbidding results, Friedman [14] noted that the costs of upward and downward deviations were approximately symmetric, so the weakness of subjects' incentives does not in itself explain *over*bidding: explaining systematic deviations from equilibrium requires asymmetric costs.

Following Friedman's suggestion, Goeree *et al.* compare two otherwise identical treatments in which the costs of deviating are strongly asymmetric, in different directions. As in previous experiments, both treatments yield overbidding, and there is a systematic difference across them in the expected direction. To sort out the effects of risk aversion and noise in subjects' bidding behavior, Goeree *et al.* use the data to estimate a model based on McKelvey and Palfrey's [27] normal-form quantal response equilibrium, generalized to allow constant relative risk aversion.[22] For a

[22] Quantal response equilibrium was proposed as a way to adapt Nash equilibrium to accommodate noisiness of players' strategy choices while preserving much of the parsimony of equilibrium analysis. Players' strategies are assumed to follow a distribution, usually characterized by a single noise parameter, in which strategies with higher expected payoffs have higher probabilities. A quantal response equilibrium is a fixed point in the space of distributions, in which players take the noise in each other's strategies rationally into account when evaluating expected payoffs.

given distribution (in this case the power distribution) and given values of the noise and risk parameters, quantal response equilibrium makes specific, probabilistic predictions about how the costs of deviating from equilibrium affect players' strategies. Econometric estimates of the resulting two-parameter model yield a sensible, unified explanation of the observed bidding behavior and the difference across treatments, which outperforms alternative explanations of overbidding.

REFERENCES

1. D. Abreu and H. Matsushima, Virtual implementation in iteratively undominated strategies: Complete information, *Econometrica* **60** (1992), 993–1008.
2. K. Binmore, J. McCarthy, G. Ponti, L. Samuelson, and A. Shaked, A backward induction experiment, *J. Econ. Theory* **104** (2002), 48–88.
3. C. Camerer, Progress in behavioral game theory, *Journal of Economic Perspectives* **4** (1997), 167–188.
4. C. Camerer, "Behavioral Game Theory: Experiments on Strategic Interaction," Princeton Univ. Press, Princeton, NJ, 2002.
5. C. Camerer and T.-H. Ho, Experience-weighted attraction learning in normal form games, *Econometrica* **67** (1999), 827–874.
6. C. F. Camerer, T.-H. Ho, and J.-K. Chong, Sophisticated experience-weighted attraction learning and strategic teaching in repeated games, *J. Econ. Theory* **104** (2002), 137–188.
7. C. Camerer, E. Johnson, T. Rymon, and S. Sen, Cognition and framing in sequential bargaining for gains and losses, *in* "Frontiers of Game Theory" (K. Binmore, A. Kirman, and P. Tani, Eds.), pp. 27–47, MIT Press, Cambridge, MA, 1993.
8. C. Camerer and K. Weigelt, Experimental tests of a sequential equilibrium reputation model, *Econometrica* **56** (1988), 1–36.
9. R. Cooper, D. DeJong, R. Forsythe, and T. Ross, Communication in the Battle of the Sexes game: Some experimental results, *Rand Journal of Economics* **20** (1989), 568–587.
10. M. A. Costa-Gomes, A suggested interpretation of some experimental results on preplay communication, *J. Econ. Theory* **104** (2002), 104–136.
11. M. Costa-Gomes, Vincent P. Crawford, and Bruno Broseta, Cognition and behavior in normal-form games: An experimental study, *Econometrica* **69** (2001), 1193–1235.
12. V. P. Crawford, Theory and experiment in the analysis of strategic interaction, *in* "Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress" (D. Kreps and K. Wallis, Eds.), Econometric Society Monograph, No. 27, Vol. I, pp. 206–242, Cambridge Univ. Press, Cambridge, U.K./New York, 1997; to be reprinted with additions *in* "Readings in Behavioral Economics" (C. Camerer, G. Loewenstein, and M. Rabin, Eds.), in press.
13. V. P. Crawford, A survey of experiments on communication via cheap talk, *J. Econ. Theory* **78** (1998), 286–298.
14. D. Friedman, Theory and misbehavior of first-price auctions: Comment, *Amer. Econ. Rev.* **82** (1992), 1374–1378.
15. D. Fudenberg and D. Levine, "The Theory of Learning in Games," MIT Press, Cambridge, MA, 1998.

16. J. Glazer and M. Perry, Virtual implementation in backwards induction, *Games Econ. Behav.* **15** (1996), 27–32.
17. J. Glazer and R. Rosenthal, A note on Abreu–Matsushima mechanisms, *Econometrica* **60** (1992), 1435–1438.
18. J. K. Goeree, C. A. Holt, and T. R. Palfrey, Quantal response equilibrium and overbidding in private-value auctions, *J. Econ. Theory* **104** (2002), 247–272.
19. G. Harrison, Theory and misbehavior of first-price auctions, *Amer. Econ. Rev.* **79** (1989), 749–762.
20. J. Harsanyi and R. Selten, "A General Theory of Equilibrium Selection in Games," MIT Press, Cambridge, MA, 1988.
21. T.-H. Ho, C. Camerer, and K. Weigelt, Iterated dominance and iterated best response in experimental "*p*-beauty contests," *Amer. Econ. Rev.* **88** (1998), 947–969.
22. E. J. Johnson, C. Camerer, S. Sen, and T. Rymon, Detecting failures of backward induction: Monitoring information search in sequential bargaining, *J. Econ. Theory* **104** (2002), 16–47.
23. J. Kagel and A. Roth, Eds., "Handbook of Experimental Economics," Princeton Univ. Press, Princeton, NJ, 1995.
24. M. Kandori, Social norms and community enforcement, *Rev. Econ. Stud.* **59** (1992), 63–80.
25. E. Katok, M. Sefton, and A. Yavas, Implementation by iterative dominance and backward induction: An experimental comparison, *J. Econ. Theory* **104** (2002), 89–103.
26. J. Ledyard and T. Palfrey, Introduction, *Games Econ. Behav.* **10** (1995), 1–5.
27. R. McKelvey and T. Palfrey, Quantal response equilibria for normal-form games, *Games Econ. Behav.* **7** (1995), 6–38.
28. R. McKelvey and T. Palfrey, Quantal response equilibria for extensive-form games, *Exper. Econ.* **1** (1998), 9–41.
29. B. O'Neill, Nonmetric test of the minimax theory of two-person zero-sum games, *Proc Nat. Acad. Sci. USA* **84** (1987), 2106–2109.
30. C. Plott, Economics in 2090: The views of an experimentalist, *Econ. J.* **101** (1991), 88–93.
31. M. Rabin, A model of pre-game communication, *J. Econ. Theory* **63** (1994), 370–391.
32. F. Rankin, J. Van Huyck, and R. Battalio, Strategic similarity and emergent conventions: Evidence from similar Stag Hunt games, *Games Econ. Behav.* **32** (2000), 315–337.
33. A. Roth, Bargaining phenomena and bargaining theory, *in* "Laboratory Experimentation in Economics: Six Points of View" (A. Roth, Ed.), pp. 14–41, Cambridge Univ. Press, Cambridge, UK/New York, 1987.
34. A. Roth, Laboratory experimentation in economics: A methodological overview, *Econ. J.* **98** (1988), 974–1031.
35. A. Roth, Game theory as a part of empirical economics, *Econ. J.* **101** (1991), 107–114.
36. L. Samuelson, Analogies, adaptation, and anomalies, *J. Econ. Theory* **97** (2001), 320–366.
37. T. Schelling, "The Strategy of Conflict," Harvard Univ. Press, Cambridge, MA, 1960.
38. M. Sefton and A. Yavas, Abreu–Matsushima mechanisms: Experimental evidence, *Games Econ. Behav.* **16** (1996), 280–302
39. R. Selten, Features of experimentally observed bounded rationality, *Eur. Econ. Rev.* **42** (1998), 413–436.
40. J. M. Shachat, Mixed strategy play and the minimax hypothesis, *J. Econ. Theory* **104** (2002), 189–226.
41. V. Smith, An experimental study of competitive market behavior, *J. Pol. Economy* **70** (1962), 111–137.

42. V. Smith, Microeconomic systems as an experimental science, *Amer. Econ. Rev.* **72** (1982), 923–955.
43. V. Smith, Theory, experiment and economics, *J. Econ. Perspect.* **3** (1989), 151–169.
44. J. Van Huyck and R. Battalio, Prudence, justice, benevolence, and sex: Evidence from similar bargaining games, *J. Econ. Theory* **104** (2002), 227–246.
45. J. Van Huyck, R. Battalio, and R. Beil, Tacit coordination games, strategic uncertainty, and coordination failure, *Amer. Econ. Rev.* **80** (1990), 234–248.