



## Ten Little Treasures of Game Theory and Ten Intuitive Contradictions

Jacob K. Goeree; Charles A. Holt

*The American Economic Review*, Vol. 91, No. 5. (Dec., 2001), pp. 1402-1422.

Stable URL:

<http://links.jstor.org/sici?sici=0002-8282%28200112%2991%3A5%3C1402%3ATLTOGT%3E2.0.CO%3B2-N>

*The American Economic Review* is currently published by American Economic Association.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/aea.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# Ten Little Treasures of Game Theory and Ten Intuitive Contradictions

By Jacob K. Goeree and Charles A. Holt\*

*This paper reports laboratory data for games that are played only once. These games span the standard categories: static and dynamic games with complete and incomplete information. For each game, the treasure is a treatment in which behavior conforms nicely to predictions of the Nash equilibrium or relevant refinement. In each case, however, a change in the payoff structure produces a large inconsistency between theoretical predictions and observed behavior. These contradictions are generally consistent with simple intuition based on the interaction of payoff asymmetries and noisy introspection about others' decisions. (JEL C72, C92)*

The Nash equilibrium has been the centerpiece of game theory since its introduction about 50 years ago. Along with supply and demand, the Nash equilibrium is one of the most commonly used theoretical constructs in economics, and it is increasingly being applied in other social sciences. Indeed, game theory has finally gained the central role envisioned by John von Neumann and Oscar Morgenstern, and in some areas of economics (e.g., industrial organization) virtually all recent theoretical developments are applications of game theory. The impression one gets from recent surveys and game theory textbooks is that the field has reached a comfortable maturity, with neat classifications of games and successively stronger (more “refined”) versions of the basic approach being appropriate for more complex categories of games: Nash equilibrium for static games with complete information, Bayesian Nash for static games with incomplete information, subgame perfectness for dynamic games with complete information, and some refinement of the

sequential Nash equilibrium for dynamic games with incomplete information (e.g., Robert Gibbons, 1997). The rationality assumptions that underlie this analysis are often preceded by persuasive adjectives like “perfect,” “intuitive,” and “divine.” If any noise in decision-making is admitted, it is eliminated in the limit in a process of “purification.” It is hard not to notice parallels with theology, and the highly mathematical nature of the developments makes this work about as inaccessible to mainstream economists as medieval treatises on theology would have been to the general public.

The discordant note in this view of game theory has come primarily from laboratory experiments, but the prevailing opinion among game theorists seems to be that behavior will eventually converge to Nash predictions under the right conditions.<sup>1</sup> This paper presents a much more unsettled perspective of the current state of game theory. In each of the major types of games, we present one or more examples for which the relevant version of the Nash equilibrium predicts remarkably well. These “treasures” are observed in games played only once by financially motivated subjects who have had prior experience in other, similar, strategic situations. In each of these games, however, we show that a change in the payoff structure can produce a large inconsistency between theoret-

\* Goeree: Department of Economics, 114 Rouss Hall, University of Virginia, P.O. Box 400182, Charlottesville, VA 22904, and University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (e-mail: jg2n@virginia.edu); Holt: Department of Economics, 114 Rouss Hall, University of Virginia, P.O. Box 400182, Charlottesville, VA 22904 (e-mail: cah2k@virginia.edu). We wish to thank Rachel Parkin and Scott Saiers for research assistance, and Colin Camerer, Monica Capra, Glenn Harrison, Susan Laury, Melayne McInnes, Theo Offerman, Amnon Rapoport, Joel Sobel, and an anonymous referee for helpful comments. This research was funded in part by the National Science Foundation (SES-9818683).

<sup>1</sup> For example, George J. Mailath's (1998) survey of evolutionary models cites the failure of backward induction as the main cause of behavioral deviations from Nash predictions.

ical prediction(s) and human behavior. For example, a payoff change that does not alter the unique Nash equilibrium may move the data to the opposite side of the range of feasible decisions. Alternatively, a payoff change may cause a major shift in the game-theoretic predictions and have no noticeable effect on actual behavior. The observed contradictions are typically somewhat intuitive, even though they are not explained by standard game theory. In a simultaneous effort-choice coordination game, for example, an increase in the cost of players' "effort" decisions is shown to cause a dramatic decrease in effort, despite the fact that *any* common effort is a Nash equilibrium for a range of effort costs. In some of these games, it seems like the Nash equilibrium works only by coincidence, e.g., in symmetric cases where the costs of errors in each direction are balanced. In other cases, the Nash equilibrium has considerable drawing power, but economically significant deviations remain to be explained.

The idea that game theory should be tested with laboratory experiments is as old as the notion of a Nash equilibrium, and indeed, the classic prisoner's dilemma paradigm was inspired by an experiment conducted at the RAND Corporation in 1950. Some of the strategic analysts at RAND were dissatisfied with the received theory of cooperative and zero-sum games in von Neumann and Morgenstern's (1944) *Theory of Games and Economic Behavior*. In particular, nuclear conflict was not thought of as a zero-sum game because both parties may lose. Sylvia Nasar (1998) describes the interest at RAND when word spread that a Princeton graduate student, John Nash, had generalized von Neumann's existence proof for zero-sum games to the class of all games with finite numbers of strategies. Two mathematicians, Melvin Dresher and Merrill Flood, had been running some game experiments with their colleagues, and they were skeptical that human behavior would be consistent with Nash's notion of equilibrium. In fact, they designed an experiment that was run on the same day they heard about Nash's proof. Each player in this game had a dominant strategy to defect, but both would earn more if they both used the cooperative strategy. The game was repeated 100 times with the same two players, and a fair amount of cooperation was observed. One of Nash's professors, Al W. Tucker, saw the pay-

offs for this game written on a blackboard, and he invented the prisoner's dilemma story that was later used in a lecture on game theory that he gave in the Psychology Department at Stanford (Tucker, 1950).

Interestingly, Nash's response to Dresher and Flood's repeated prisoner's dilemma experiment is contained in a note to the authors that was published as a footnote to their paper:

The flaw in the experiment as a test of equilibrium point theory is that the experiment really amounts to having the players play one large multi-move game. One cannot just as well think of the thing as a sequence of independent games as one can in zero-sum cases. There is just too much interaction . . . (Nasar, 1998 p. 119).

In contrast, the experiments that we report in this paper involved games that were played *only once*, although related results for repeated games with random matching will be cited where appropriate. As Nash noted, the advantage of one-shot games is that they insulate behavior from the incentives for cooperation and reciprocity that are present in repeated games. One potential disadvantage of one-shot games is that, without an opportunity to learn and adapt, subjects may be especially prone to the effects of confusion. The games used in this paper, however, are simple enough in structure to ensure that Nash-like behavior can be observed in the "treasure" treatment. In addition, the study of games played only once is of independent interest given the widespread applications of game theory to model one-shot interactions in economics and other social sciences, e.g., the FCC license auctions, elections, military campaigns, and legal disputes.

The categories of games to be considered are based on the usual distinctions: static versus dynamic and complete versus incomplete information. Section I describes the experiments based on static games with complete information: social dilemma, matching pennies, and coordination games. Section II contains results from dynamic games with complete information: bargaining games and games with threats that are not credible. The games reported in Sections III and IV have incomplete information about other players' payoffs: in static

settings (auctions) and two-stage settings (signaling games).

It is well known that decisions can be affected by psychological factors such as framing, aspiration levels, social distance, and heuristics (e.g., Daniel Kahneman et al., 1982; Catherine Eckel and Rick Wilson, 1999). In this paper we try to hold psychological factors constant and focus on payoff changes that are primarily *economic* in nature. As noted below, economic theories can and are being devised to explain the resulting anomalies. For example, the rational-choice assumption underlying the notion of a Nash equilibrium eliminates all errors, but if the costs of “overshooting” an optimal decision are much lower than the costs of “undershooting,” one might expect an upward bias in decisions. In a game, the endogenous effects of such biases may be reinforcing in a way that creates a “snowball” effect that moves decisions well away from a Nash prediction. Models that introduce (possibly small) amounts of noise into the decision-making process can produce predictions that are quite far from any Nash equilibrium (Richard D. McKelvey and Thomas R. Palfrey, 1995, 1998; Goeree and Holt, 1999). Equilibrium models of noisy behavior have been used to explain behavior in a variety of contexts, including jury decision-making, bargaining, public goods games, imperfect price competition, and coordination (Simon P. Anderson et al., 1998a, b, 2001a; C. Monica Capra et al., 1999, 2002; Stanley S. Reynolds, 1999; Serena Guarnaschelli et al., 2001).

A second type of rationality assumption that is built into the Nash equilibrium is that beliefs are consistent with actual decisions. Beliefs are not likely to be confirmed out of equilibrium, and learning will presumably occur in such cases. There is a large recent literature on incorporating learning into models of adjustment in games that are played repeatedly with different partners.<sup>2</sup> These models include adaptive learning (e.g., Vincent P. Crawford, 1995; David J. Cooper et al., 1997), naive Bayesian learning (e.g., Jordi Brandts and Holt, 1996; Dilip Mookherjee and Barry Sopher, 1997), reinforcement or stimulus-response learning (e.g., Ido Erev and Alvin E. Roth, 1998), and hybrid

models with elements of both belief and reinforcement learning (Colin Camerer and Teck-Hua Ho, 1999). Learning from experience is not possible in games that are only played once, and beliefs must be formed from introspective thought processes, which may be subject to considerable noise. Without noise, iterated best responses will converge to a Nash equilibrium, if they converge at all. Some promising approaches to explaining deviations from Nash predictions are based on models that limit players’ capacities for introspection, either by limiting the number of iterations (e.g., Dale O. Stahl and Paul W. Wilson, 1995; Rosemarie Nagel, 1995) or by injecting increasing amounts of noise into higher levels of iterated beliefs (Goeree and Holt, 1999; Dorothea Kübler and Georg Weizsäcker, 2000). The predictions derived from these approaches, discussed in Section V, generally conform to Nash predictions in the treasure treatments and to the systematic, intuitive deviations in the contradiction treatments. Some conclusions are offered in Section VI.

### I. Static Games with Complete Information

In this section we consider a series of two-player, simultaneous-move games, for which the Nash equilibria show an increasing degree of complexity. The first game is a “social dilemma” in which the pure-strategy Nash equilibrium coincides with the unique rationalizable outcome. Next, we consider a matching pennies game with a unique Nash equilibrium in mixed strategies. Finally, we discuss coordination games that have multiple Nash equilibria, some of which are better for all players.

In all of the games reported here and in subsequent sections, we used cohorts of student subjects recruited from undergraduate economics classes at the University of Virginia. Each cohort consisted of ten students who were paid \$6 for arriving on time, plus all cash they earned in the games played. These one-shot games followed an initial “part A” in which the subjects played the same two-person game for ten periods with new pairings made randomly in each period.<sup>3</sup> Earnings for the two-hour ses-

<sup>2</sup> See, for instance, Drew Fudenberg and David K. Levine (1998) for a survey.

<sup>3</sup> We only had time to run about six one-shot games in each session, so the data are obtained from a large number

sions ranged from \$15 to \$60, with an average of about \$35. Each one-shot game began with the distribution and reading of the instructions for that game.<sup>4</sup> These instructions contained assurances that all money earned would be paid and that the game would be followed by “another, quite different, decision-making experiment.” Since the one-shot treatments were paired, we switched the order of the treasure and contradiction conditions in each subsequent session. Finally, the paired treatments were always separated by other one-shot games of a different type.

#### A. The One-Shot Traveler’s Dilemma Game

The Nash equilibrium concept is based on the twin assumptions of perfect error-free decision-making and the consistency of actions and beliefs. The latter requirement may seem especially strong in the presence of multiple equilibria when there is no obvious way for players to coordinate. More compelling arguments can be given for the Nash equilibrium when it predicts the play of the unique justifiable, or *rationalizable*, action (B. Douglas Bernheim, 1984; David G. Pearce, 1984). Rationalizability is based on the idea that players should eliminate those strategies that are never a best response for any possible beliefs, and realize that other (rational) players will do the same.<sup>5</sup>

To illustrate this procedure, consider the game in which two players independently and simultaneously choose integer numbers between (and including) 180 and 300. Both players are paid the *lower* of the two numbers, and, in addition, an amount  $R > 1$  is transferred from the player with the higher number to the player with the lower number. For instance, if one person chooses 210 and the other chooses 250, they receive payoffs of  $210 + R$  and

$210 - R$  respectively. Since  $R > 1$ , the best response is to undercut the other’s decision by 1 (if that decision were known), and therefore, the upper bound 300 is never a best response to any possible beliefs that one could have. Consequently, a rational person must assign a probability of zero to a choice of 300, and hence 299 cannot be a best response to any possible beliefs that rule out choices of 300, etc. Only the lower bound 180 survives this iterated deletion process and is thus the unique rationalizable action, and hence the unique Nash equilibrium.<sup>6</sup> This game was introduced by Kaushik Basu (1994) who coined it the “traveler’s dilemma” game.<sup>7</sup>

Although the Nash equilibrium for this game can be motivated by successively dropping those strategies that are never a best response (to any beliefs about strategies that have not yet been eliminated from consideration), this deletion process may be too lengthy for human subjects with limited cognitive abilities. When the cost of having the higher number is small, i.e., for small values of  $R$ , one might expect more errors in the direction of high claims, well away from the unique equilibrium at 180, and indeed this is the intuition behind the dilemma. In contrast, with a large penalty for having the higher of the two claims, players are likely to end up with claims that are near the unique Nash prediction of 180.

To test these hypotheses we asked 50 subjects (25 pairs) to make choices in a treatment with  $R = 180$ , and again in a matched treatment with  $R = 5$ . All subjects made decisions in each treatment, and the two games were separated by a number of other one-shot games. The

<sup>6</sup> In other games, rationalizability may allow outcomes that are not Nash equilibria, so it is a weaker concept than that of a Nash equilibrium, allowing a wider range of possible behavior. It is in this sense that Nash is more persuasive when it corresponds to the unique rationalizable outcome.

<sup>7</sup> The associated story is that two travelers purchase identical antiques while on a tropical vacation. Their luggage is lost on the return trip, and the airline asks them to make independent claims for compensation. In anticipation of excessive claims, the airline representative announces: “We know that the bags have identical contents, and we will entertain any claim between \$180 and \$300, but you will each be reimbursed at an amount that equals the *minimum* of the two claims submitted. If the two claims differ, we will also pay a reward  $R$  to the person making the smaller claim and we will deduct a penalty  $R$  from the reimbursement to the person making the larger claim.”

of sessions where part A involved a wide range of repeated games, including public goods, coordination, price competition, and auction games that are reported in other papers. The one-shot games never followed a repeated game of the same type.

<sup>4</sup> These instructions can be downloaded from <http://www.people.virginia.edu/~cah2k/datapage.html>.

<sup>5</sup> A well-known example for which this iterated deletion process results in a unique outcome is a Cournot duopoly game with linear demand (Fudenberg and Jean Tirole, 1993 pp. 47–48).

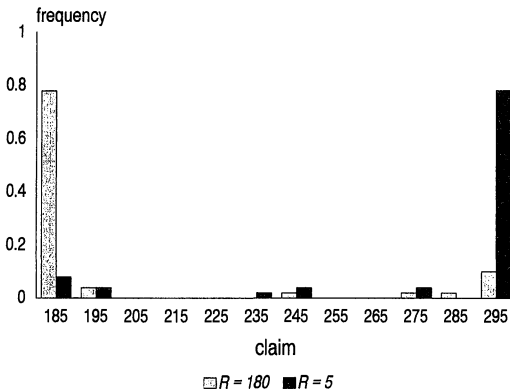


FIGURE 1. CLAIM FREQUENCIES IN A TRAVELER'S DILEMMA FOR  $R = 180$  (LIGHT BARS) AND  $R = 5$  (DARK BARS)

ordering of the two treatments was alternated. The instructions asked the participants to devise their own numerical examples to be sure that they understood the payoff structure.

Figure 1 shows the frequencies for each 10-percent category centered around the claim label on the horizontal axis. The lighter bars pertain to the high- $R$  "treasure" treatment, where close to 80 percent of all the subjects chose the Nash equilibrium strategy, with an average claim of 201. However, roughly the same fraction chose the *highest* possible claim in the low- $R$  treatment, for which the average was 280, as shown by the darker bars. Notice that the data in the contradiction treatment are clustered at the opposite end of the set of feasible decisions from the unique (rationalizable) Nash equilibrium.<sup>8</sup> Moreover, the "anomalous" result for the low- $R$  treatment does not disappear or even diminish over time when subjects play the game repeatedly and have the opportunity to learn.<sup>9</sup> Since

<sup>8</sup> This result is statistically significant at all conventional levels, given the strong treatment effect and the relatively large number of independent observations (two paired observations for each of the 50 subjects). We will not report specific nonparametric tests for cases that are so clearly significant. The individual choice data are provided in the Data Appendix for this paper on: <http://www.people.virginia.edu/~cah2k/datapage.html>.

<sup>9</sup> In Capra et al. (1999), we report results of a repeated traveler's dilemma game (with random matching). When subjects chose numbers in the range [80, 200] with  $R = 5$ , the average claim *rose* from approximately 180 in the first period to 196 in period 5, and the average remained above 190 in later periods. Different cohorts played this game with different values of  $R$ , and successive increases in  $R$  resulted

TABLE 1—THREE ONE-SHOT MATCHING PENNIES GAMES (WITH CHOICE PERCENTAGES)

		<i>Left</i> (48)	<i>Right</i> (52)
Symmetric matching pennies	<i>Top</i> (48)	80, 40	40, 80
	<i>Bottom</i> (52)	40, 80	80, 40
Asymmetric matching pennies	<i>Top</i> (96)	<i>Left</i> (16) 320, 40	<i>Right</i> (84) 40, 80
	<i>Bottom</i> (4)	40, 80	80, 40
Reversed asymmetry	<i>Top</i> (8)	<i>Left</i> (80) 44, 40	<i>Right</i> (20) 40, 80
	<i>Bottom</i> (92)	40, 80	80, 40

the treatment change does not alter the unique Nash (and rationalizable) prediction, standard game theory simply cannot explain the most salient feature of the data, i.e., the effect of the penalty/reward parameter on average claims.

### B. A Matching Pennies Game

Consider a symmetric matching pennies game in which the row player chooses between *Top* and *Bottom* and the column player simultaneously chooses between *Left* and *Right*, as shown in top part of Table 1. The payoff for the row player is \$0.80 when the outcome is (*Top*, *Left*) or (*Bottom*, *Right*) and \$0.40 otherwise. The motivations for the two players are exactly opposite: column earns \$0.80 when row earns \$0.40, and vice versa. Since the players have opposite interests there is no equilibrium in pure strategies. Moreover, in order not to be exploited by the opponent, neither player should favor one of their strategies, and the mixed-strategy Nash equilibrium involves randomizing over both alternatives with equal probabilities. As before, we obtained decisions from 50 subjects in a one-shot version of this game (five cohorts of ten subjects, who were randomly matched and assigned row or column

in successive reductions in average claims. With a penalty/reward parameter of 5, 10, 20, 25, 50, and 80 the average claims in the final three periods were 195, 186, 119, 138, 85, and 81 respectively. Even though there is one treatment reversal, the effect of the penalty/reward parameter on average claims is significant at the 1-percent level. The patterns of adjustment are well explained by a naive Bayesian learning model with decision error, and the claim distributions for the final five periods are close to those predicted by a logit equilibrium (McKelvey and Palfrey, 1995).

roles). The choice percentages are shown in parentheses next to the decision labels in the top part of Table 1. Note that the choice percentages are essentially “50-50,” or as close as possible given that there was an odd number of players in each role.

Now consider what happens if the row player’s payoff of \$0.80 in the (*Top, Left*) box is increased to \$3.20, as shown in the asymmetric matching pennies game in the middle part of Table 1. In a mixed-strategy equilibrium, a player’s *own* decision probabilities should be such that the *other* player is made indifferent between the two alternatives. Since the column player’s payoffs are unchanged, the mixed-strategy Nash equilibrium predicts that *row’s* decision probabilities do not change either. In other words, the row player should ignore the unusually high payoff of \$3.20 and still choose *Top* or *Bottom* with probabilities of one-half. (Since column’s payoffs are either 40 or 80 for playing *Left* and either 80 or 40 for playing *Right*, row’s decision probabilities must equal one-half to keep column indifferent between *Left* and *Right*, and hence willing to randomize.)<sup>10</sup> This counterintuitive prediction is dramatically rejected by the data, with 96 percent of the row players choosing the *Top* decision that gives a chance of the high \$3.20 payoff. Interestingly, the column players seemed to have anticipated this, and they played *Right* 84 percent of the time, which is quite close to their equilibrium mixed strategy of  $\frac{7}{8}$ . Next, we lowered the row player’s (*Top, Left*) payoff to \$0.44, which again should leave the row player’s own choice probabilities unaffected in a mixed-strategy Nash equilibrium. Again the effect is dramatic, with 92 percent of the choices being *Down*, as shown in the bottom part of Table 1. As before, the column players seemed to have anticipated this reaction, playing *Left* 80 percent of the time. To summarize, the unique Nash prediction is for the bolded row-choice percentages to be unchanged at 50 percent for all three treatments. This prediction is violated

<sup>10</sup> The predicted equilibrium probabilities for the row player are not affected if we relax the assumption of risk neutrality. There are only two possible payoff levels for column so, without loss of generality, columns’ utilities for payoffs of 40 and 80 can be normalized to 0 and 1. Hence even a risk-averse column player will only be indifferent when row uses choice probabilities of one-half.

TABLE 2—AN EXTENDED COORDINATION GAME

	<i>L</i>	<i>H</i>	<i>S</i>
<i>L</i>	90, 90	0, 0	$x$ , 40
<i>H</i>	0, 0	180, 180	0, 40

in an intuitive manner, with row players’ choices responding to their own payoffs.<sup>11</sup> In this context, *the Nash mixed-strategy prediction seems to work only by coincidence*, when the payoffs are symmetric.

### C. A Coordination Game with a Secure Outside Option

Games with multiple Nash equilibria pose interesting new problems for predicting behavior, especially when some equilibria produce higher payoffs for all players. The problem of coordinating on the high-payoff equilibrium may be complicated by the possible gains and losses associated with payoffs that are not part of any equilibrium outcome. Consider a coordination game in which players receive \$1.80 if they coordinate on the high-payoff equilibrium (*H, H*) \$0.90 if they coordinate on the low-payoff equilibrium (*L, L*), and they receive nothing if they fail to coordinate (i.e., when one player chooses *H* and the other *L*). Suppose that, in addition, the column player has a secure option *S* that yields \$0.40 for column and results in a zero payoff for the row player. This game is given in Table 2 when  $x = 0$ . To analyze the Nash equilibria of this game, notice that for the column player a 50-50 combination of *L* and *H* dominates *S*, and a rational column player should therefore avoid the secure option. Eliminating *S* turns the game into a standard  $2 \times 2$  coordination game that has three Nash equilibria:

<sup>11</sup> This anomaly is persistent when subjects play the game repeatedly. Jack Ochs (1995a, b) investigates a matching pennies game with an asymmetry similar to that of the middle game in Table 1, and reports that the row players continue to select *Top* considerably more than one-half of the time, even after as many as 50 rounds. These results are replicated in McKelvey et al. (2000). Similarly, Goeree et al. (2000) report results for ten-period repeated matching pennies games that exactly match those in Table 1. The results are qualitatively similar but less dramatic than those in Table 1, with row’s choice probabilities showing strong “own-payoff” effects that are not predicted by the Nash equilibrium.

both players choosing  $L$ , both choosing  $H$ , and a mixed-strategy equilibrium in which both players choose  $L$  with probability  $\frac{2}{3}$ .

The Nash equilibria are independent of  $x$ , which is the payoff to the row player when  $(L, S)$  is the outcome, since the argument for eliminating  $S$  is based solely on column's payoffs. However, the magnitude of  $x$  may affect the coordination process: for  $x = 0$ , row is indifferent between  $L$  and  $H$  when column selects  $S$ , and row is likely to prefer  $H$  when column does not select  $S$  (since then  $L$  and  $H$  have the same number of zero payoffs for row, but  $H$  has a higher potential payoff). Row is thus more likely to choose  $H$ , which is then also the optimal action for the column player. However, when  $x$  is large, say 400, the column player may anticipate that row will select  $L$  in which case column should avoid  $H$ .

This intuition is borne out by the experimental data: in the treasure treatment with  $x = 0$ , 96 percent of the row players and 84 percent of the column players chose the high-payoff action  $H$ , while in the contradiction treatment with  $x = 400$  only 64 percent of the row players and 76 percent of the column players chose  $H$ . The percentages of outcomes that were coordinated on the high-payoff equilibrium were 80 for the treasure treatment versus 32 for the contradiction treatment. In the latter treatment, an additional 16 percent of the outcomes were coordinated on the low-payoff equilibrium, but more than half of all the outcomes were uncoordinated, non-Nash outcomes.

D. A Minimum-Effort Coordination Game

The next game we consider is also a coordination game with multiple equilibria, but in this case the focus is on the effect of payoff asymmetries that determine the risks of deviating in the upward and downward directions. The two players in this game choose "effort" levels simultaneously, and the cost of effort determines the risk of deviation. The joint product is of the fixed-coefficients variety, so that each person's payoff is the minimum of the two efforts, minus the product of the player's own effort and a constant cost factor,  $c$ . In the experiment, we let efforts be any integer in the range from 110 to 170. If  $c < 1$ , any common effort in this range is a Nash equilibrium, because a unilateral one-unit increase in effort above a

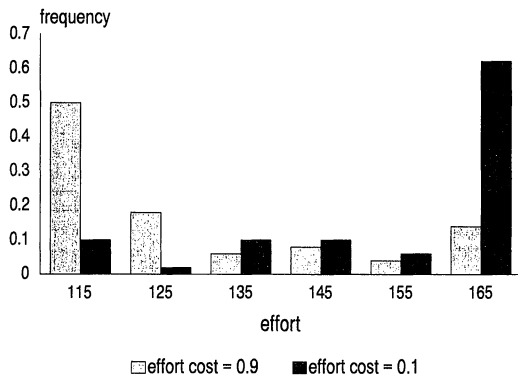


FIGURE 2. EFFORT CHOICE FREQUENCIES FOR A MINIMUM-EFFORT COORDINATION GAME WITH HIGH EFFORT COST (LIGHT BARS) AND LOW EFFORT COST (DARK BARS)

common starting point will not change the minimum but will reduce one's payoff by  $c$ . Similarly, a one-unit decrease in effort will reduce one's payoff by  $1 - c$ , i.e., the reduction in the minimum is more than the savings in effort costs when  $c < 1$ . Obviously, a higher effort cost increases the risk of raising effort and reduces the risk of lowering effort. Thus simple intuition suggests that effort levels will be inversely related to effort costs, despite the fact that any common effort level is a Nash equilibrium.

We ran one treatment with a low effort cost of 0.1, and the data for 50 randomly matched subjects in this treatment are shown by the dark bars in Figure 2. Notice that behavior is quite concentrated at the highest effort level of 170; subjects coordinate on the Pareto-dominant outcome. The high effort cost treatment ( $c = 0.9$ ), however, produced a preponderance of efforts at the lowest possible level, as can be seen by the lighter bars in the figure. Clearly, the extent of this "coordination failure" is affected by the key economic variable in this model, even though Nash theory is silent.<sup>12</sup>

<sup>12</sup> The standard analysis of equilibrium selection in coordination is based on the John C. Harsanyi and Reinhard Selten's (1988) notion of risk dominance, which allows a formal analysis of the trade-off between risk and payoff dominance. Paul G. Straub (1995) reports experimental evidence for risk dominance as a selection criterion. There is no agreement on how to generalize risk dominance beyond  $2 \times 2$  games, but see Anderson et al. (2001b) for a proposed generalization based on the "stochastic potential." Experiments with repeated plays of coordination games



TABLE 3—TWO VERSIONS OF THE KREPS GAME (WITH CHOICE PERCENTAGES)

		<i>Left</i> (26)	<i>Middle</i> (8)	<i>Non-Nash</i> (68)	<i>Right</i> (0)
Basic game	<i>Top</i> (68)	200, 50	0, 45	10, 30	20, -250
	<i>Bottom</i> (32)	0, -250	10, -100	30, 30	50, 40
Positive payoff frame	<i>Top</i> (84)	500, 350	300, 345	310, 330	320, 50
	<i>Bottom</i> (16)	300, 50	310, 200	330, 330	350, 340

### E. *The Kreps Game*

The previous examples demonstrate how the cold logic of game theory can be at odds with intuitive notions about human behavior. This tension has not gone unnoticed by some game theorists. For instance, David M. Kreps (1995) discusses a variant of the game in the top part of Table 3 (where we have scaled back the payoffs to levels that are appropriate for the laboratory). The pure-strategy equilibrium outcomes of this game are (*Top, Left*) and (*Bottom, Right*). In addition, there is a mixed-strategy equilibrium in which row randomizes between *Top* and *Bottom* and column randomizes between *Left* and *Middle*. The only column strategy that is not part of any Nash equilibrium is labeled *Non-Nash*. Kreps argues that column players will tend to yield *Non-Nash* because the other options yield at best a slightly higher payoff (i.e., 10, 15, or 20 cents higher) but could lead to substantial losses of \$1 or \$2.50. Notice that this intuition is based on payoff magnitudes out of equilibrium, in contrast to Nash calculations based only on signs of payoff differences.

Kreps did try the high-hypothetical-payoff version of this game on several graduate students, but let us consider what happens with financially motivated subjects in an anonymous

laboratory situation. As before, we randomly paired 50 subjects and let them make a single choice. Subjects were told that losses would be subtracted from prior earnings, which were quite substantial by that point. As seen from the percentages in parentheses in the top part of the table, the *Non-Nash* decision was selected by approximately two-thirds of the column players. Of course, it is possible that this result is simply a consequence of “loss-aversion,” i.e., the disutility of losing some amount of money is greater than the utility associated with winning the same amount (Daniel Kahneman et al., 1991). Since all the other columns contain negative payoffs, loss-averse subjects would thus be naturally inclined to choose *Non-Nash*. Therefore, we ran another 50 subjects through the same game, but with 300 cents added to payoffs to avoid losses, as shown in the bottom part of Table 3. The choice percentages shown in parentheses indicate very little change, with close to two-thirds of column players choosing *Non-Nash* as before. Thus “loss aversion” biases are not apparent in the data, and do not seem to be the source of the prevalence of *Non-Nash* decisions. Finally, we ran 50 new subjects through the original version in the top part of the table, with the (*Bottom, Right*) payoffs of (50, 40) being replaced by (350, 400), which (again) does not alter the equilibrium structure of the game. With this admittedly heavy-handed enhancement of the equilibrium in that cell, we observed 96 percent *Bottom* choices and 84 percent *Right* choices, with 16 percent *Non-Nash* persisting in this, the “treasure” treatment.

## II. Dynamic Games with Complete Information

As game theory became more widely used in fields like industrial organization, the complexity of the applications increased to accommodate

---

have shown that behavior may begin near the Pareto-dominant equilibrium, but later converge to the equilibrium that is worst for all concerned (John B. Van Huyck et al., 1990). Moreover, the equilibrium that is selected may be affected by the payoff structure for dominated strategies (Russell Cooper et al., 1992). See Goeree and Holt (1998) for results of a repeated coordination game with random matching. They show that the dynamic patterns of effort choices are well explained by a simple evolutionary model of noisy adjustment toward higher payoffs, and that final-period effort decisions can be explained by the maximization of stochastic potential function.

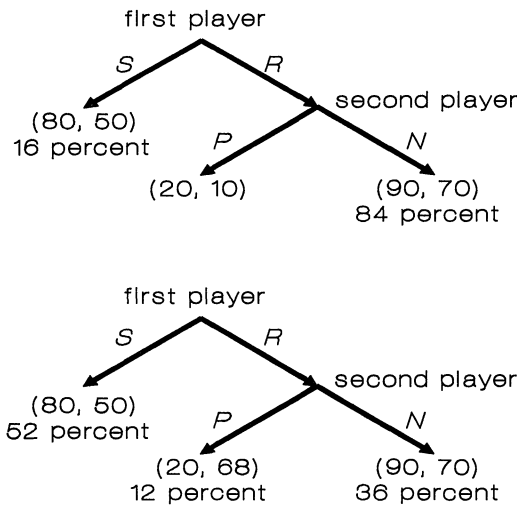


FIGURE 3. SHOULD YOU TRUST OTHERS TO BE RATIONAL?

dynamics and asymmetric information. One of the major developments coming out of these applications was the use of backward induction to eliminate equilibria with threats that are not “credible” (Selten, 1975). Backward induction was also used to develop solutions to alternating-offer bargaining games (Ariel Rubinstein, 1982), which was the first major advance on this historically perplexing topic since Nash’s (1950) axiomatic approach. However, there have been persistent doubts that people are able to figure out complicated, multistage backward induction arguments. Robert W. Rosenthal (1981) quickly proposed a game, later dubbed the “centipede game,” in which backward induction over a large number of stages (e.g., 100 stages) was thought to be particularly problematic (e.g., McKelvey and Palfrey, 1992). Many of the games in this section are inspired by Rosenthal’s (1981) doubts and Randolph T. Beard and Beil’s (1994) experimental results. Indeed, the anomalies in this section are better known than those in other sections, but we focus on very simple games with two or three stages, using parallel procedures and subjects who have previously made a number of strategic decisions in different one-shot games.

#### A. Should You Trust Others to Be Rational?

The power of backward induction is illustrated in the top game in Figure 3. The first

player begins by choosing between a safe decision, *S*, and a risky decision, *R*. If *R* is chosen, the second player must choose between a decision *P* that punishes both of them and a decision *N* that leads to a Nash equilibrium that is also a joint-payoff maximum. There is, however, a second Nash equilibrium where the first player chooses *S* and the second chooses *P*. The second player has no incentive to deviate from this equilibrium because the self-inflicted punishment occurs off of the equilibrium path. Subgame perfectness rules out this equilibrium by requiring equilibrium behavior in each subgame, i.e., that the second player behave optimally in the event that the second-stage subgame is reached.

Again, we used 50 randomly paired subjects who played this game only once. The data for this treasure treatment are quite consistent with the subgame-perfect equilibrium; a preponderance of first players trust the other’s rationality enough to choose *R*, and there are no irrational *P* decisions that follow. The game shown in the bottom part of Figure 3 is identical, except that the second player only forgoes 2 cents by choosing *P*. This change does not alter the fact that there are two Nash equilibria, one of which is ruled out by subgame perfectness. The choice percentages for 50 subjects indicate that a majority of the first players did not trust others to be perfectly rational when the cost of irrationality is so small. Only about a third of the outcomes matched the subgame-perfect equilibrium in this game.<sup>13</sup> We did a third treatment (not shown) in which we multiplied all payoffs by a factor of 5, except that the *P* decision led to (100, 348) instead of (100, 340). This large increase in payoffs produced an even more dramatic result; only 16 percent of the outcomes were subgame perfect, and 80 percent of the outcomes were at the Nash equilibrium that is not subgame perfect.

#### B. Should You Believe a Threat That Is Not Credible?

The game just considered is a little unusual in that, in the absence of relative payoff effects, the second player has no reason to punish, since

<sup>13</sup> See Beard and Beil (1994) for similar results in a two-stage game played only once.

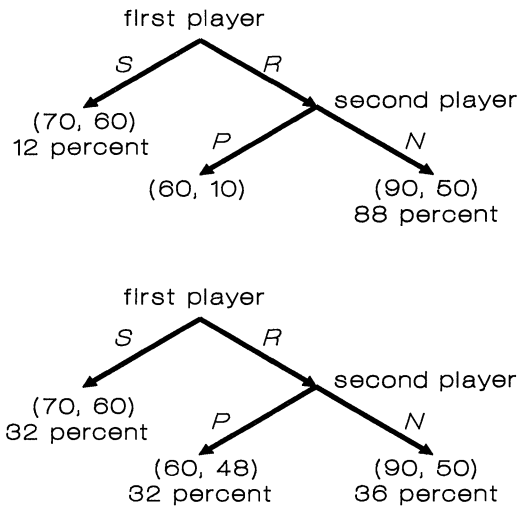


FIGURE 4. SHOULD YOU BELIEVE A THREAT THAT IS NOT CREDIBLE?

the first player's *R* decision also benefits the second player. This is not the case for the game in Figure 4, where an *R* decision by the first player will lower the second player's payoff. As before, there are two Nash equilibria, with the (*R*, *P*) equilibrium ruled out by subgame perfectness. In addition to not being credible, the threat to play *P* is a relatively costly punishment for the second player to administer (40 cents).

The threat to play *P* in the top part of Figure 4 is evidently not believed, and 88 percent of the first players choose the *R* strategy, with impunity. The threat is cheap (2 cents) for the game in the bottom part of the figure, and outcomes for 25 subject pairs are evenly divided between the subgame-imperfect outcome, the incredible threat outcome, and the subgame-perfect outcome. Cheap threats often are (and apparently should be) believed. Again we see that payoff magnitudes and off-the-equilibrium-path risks matter.

Since the *P* decisions in the bottom games of Figures 3 and 4 only reduce the second player's payoff by 2 cents, behavior may be affected by small variations in payoff preferences or emotions, e.g., spite or rivalry. As suggested by Ernst Fehr and Klaus Schmidt (1999) and Gary E Bolton and Axel Ockenfels (2000), players may be willing to sacrifice own earnings in order to reduce payoff inequities which would explain the *P* choices in the contradiction treat-

ments. Alternatively, the occurrence of the high fraction of *P* decisions in the bottom game of Figure 4 may be due to negative emotions that follow the first player's *R* decision, which reduces the second player's earnings (Matthew Rabin, 1993). Notice that this earnings reduction does not occur when the first player chooses *R* for the game in the bottom part of Figure 3, which could explain the lower rate of punishments in that game.

The anomalous results of the contradiction treatments may not come as any surprise to Selten, the originator of the notion of subgame perfectness. His attitude toward game theory has been that there is a sharp contrast between standard theory and behavior. For a long time he essentially wore different hats when he did theory and ran experiments, although his 1994 Nobel prize was clearly for his contributions in theory. This schizophrenic stance may seem inconsistent, but it may prevent unnecessary anxiety, and some of Selten's recent theoretical work is based on models of boundedly rational (directional) learning (Selten and Joachim Buchta, 1998). In contrast, John Nash was reportedly discouraged by the predictive failures of game theory and gave up on both experimentation and game theory (Nasar, 1998 p. 150).

### C. Two-Stage Bargaining Games

Bargaining has long been considered a central part of economic analysis, and at the same time, one of the most difficult problems for economic theory. One promising approach is to model unstructured bargaining situations "as if" the parties take turns making offers, with the costs of delayed agreement reflected in a shrinking size of the pie to be divided. This problem is particularly easy to analyze when the number of alternating offers is fixed and small.

Consider a bargaining game in which each player gets to make a single proposal for how to split a pie, but the amount of money to be divided falls from \$5 in the first stage to \$2 in the second. The first player proposes a split of \$5 that is either accepted (and implemented) or rejected, in which case the second player proposes a split of \$2 that is either accepted or rejected by the first player. This final rejection results in payoffs of zero for both players, so the second player can (in theory) successfully demand \$1.99 in the second stage if the first player

prefers a penny to nothing. Knowing this, the first player should demand \$3 and offer \$2 to the other in the first stage. In a subgame-perfect equilibrium, the first player receives the amount by which the pie shrinks, so a larger cost of delay confers a greater advantage to the player making the initial demand, which seems reasonable. For example, a similar argument shows that if the pie shrinks by \$4.50, from \$5 to \$0.50, then the first player should make an initial demand of \$4.50.

We used 60 subjects (six cohorts of ten subjects each), who were randomly paired for each of the two treatments described above (alternating in order and separated by other one-shot games). The average demand for the first player was \$2.83 for the \$5/\$2 treatment, with a standard deviation of \$0.29. This is quite close to the predicted \$3.00 demand, and 14 of the 30 initial demands were exactly equal to \$3.00 in this treasure treatment. But the average demand only increased to \$3.38 for the other treatment with a \$4.50 prediction, and 28 of the 30 demands were below the prediction of \$4.50. Rejections were quite common in this contradiction treatment with higher demands and correspondingly lower offers to the second player, which is not surprising given the smaller costs of rejecting “stingy” offers.

These results fit into a larger pattern surveyed in Douglas D. Davis and Holt (1993 Chapter 5) and Roth (1995); initial demands in two-stage bargaining games tend to be “too low” relative to theoretical predictions when the equilibrium demand is high, say more than 80 percent of the pie as in our \$5.00/\$0.50 treatment, and initial demands tend to be close to predictions when the equilibrium demand is 50–75 percent of the pie (as in our \$5.00/\$2.00 treatment). Interestingly, initial demands are “too high” when the equilibrium demand is less than half of the pie. Here is an example of why theoretical explanations of behavior should not be based on experiments in only one part of the parameter space, and why theorists should have more than just a casual, secondhand knowledge of the experimental economics literature.<sup>14</sup> Many of the diverse theoretical explanations for anomalous

behavior in bargaining games hinge on models of preferences in which a person’s utility depends on the payoffs of both players, i.e., distribution matters (Bolton, 1998; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Miguel Costa-Gomes and Klaus G. Zauner, 2001). The role of fairness is illustrated dramatically in the experiment reported in Goeree and Holt (2000a), who obtained even larger deviations from subgame-perfect Nash predictions than those reported here by giving subjects asymmetric money endowments that were paid independently of the bargaining outcome. These endowments were selected to accentuate the payoff inequities that result in the subgame-perfect Nash equilibrium, and hence their effect was to exaggerate fairness issues without altering the equilibrium prediction. The result (for seven different one-shot bargaining games) was for demands to be *inversely* related to the subgame-perfect Nash predictions.

### III. Static Games with Incomplete Information

William Vickrey’s (1961) models of auctions with incomplete information constitute one of the most widely used applications of game theory. If private values are drawn from a uniform distribution, the Bayesian Nash equilibrium predicts that bids will be proportional to value, which is generally consistent with laboratory evidence. The main deviation from theoretical predictions is the tendency of human subjects to “overbid” (relative to Nash), which is commonly rationalized in terms of risk aversion, an explanation that has led to some controversy. Glenn W. Harrison (1989), for instance, argues that deviations from the Nash equilibrium may well be caused by a lack of monetary incentives since the costs of such deviations are rather small: the “flat maximum critique.” Our approach here is to specify two auction games with identical Nash equilibria, but with differing incentives not to overbid.

First, consider a game in which each of two bidders receives a private value for a prize to be auctioned in a first-price, sealed-bid auction. In other words, the prize goes to the highest bidder for a price equal to that bidder’s own bid. Each bidder’s value for the prize is equally likely to be \$0, \$2, or \$5. Bids are constrained to be integer dollar amounts, with ties decided by the flip of a coin.

The relevant Nash equilibrium in this game

<sup>14</sup> Another example is the development of theories of generalized expected utility to explain “fanning out” preferences in Allais paradox situations, when later experiments in other parts of the probability triangle found “fanning in.”

TABLE 4—EQUILIBRIUM EXPECTED PAYOFFS FOR THE (0,2,5) TREATMENT (EQUILIBRIUM BIDS MARKED WITH AN ASTERISK\*)

	Bid = 0	Bid = 1	Bid = 2	Bid = 3	Bid = 4	Bid = 5
Value = \$0	0*	-0.5	-1.66	-3	-4	-5
Value = \$2	0.33	0.5*	0	-1	-2	-3
Value = \$5	0.83	2	2.5*	2	1	0

TABLE 5—EQUILIBRIUM EXPECTED PAYOFFS FOR THE (0,3,6) TREATMENT (EQUILIBRIUM BIDS MARKED WITH AN ASTERISK\*)

	Bid = 0	Bid = 1	Bid = 2	Bid = 3	Bid = 4	Bid = 5
Value = \$0	0*	-0.5	-1.66	-3	-4	-5
Value = \$3	0.5	1*	0.83	0	-1	-2
Value = \$6	1	2.5	3.33*	3	2	1

with incomplete information about others' preferences is the Bayesian Nash equilibrium, which specifies an equilibrium bid for each possible realization of a bidder's value. It is straightforward but tedious to verify that the Nash equilibrium bids are \$0, \$1, and \$2 for a value of \$0, \$2, and \$5 respectively, as can be seen from the equilibrium expected payoffs in Table 4. For example, consider a bidder with a private value of \$5 (in the bottom row) who faces a rival that bids according to the proposed Nash solution. A bid of 0 has a one-half chance of winning (decided by a coin flip) if the rival's value, and hence the rival's bid, is zero, which happens with probability one-third. Therefore, the expected payoff of a zero bid with a value of \$5 equals  $\frac{1}{2} * \frac{1}{3} * (\$5 - \$0) = \$5/6 = 0.83$ . If the bid is raised to \$1, the probability of winning becomes  $\frac{1}{2}$  ( $\frac{1}{3}$  when the rival's value is \$0 plus  $\frac{1}{6}$  when the rival's value is \$2). Hence, the expected payoff of a \$1 bid is  $\frac{1}{2} * (\$5 - \$1) = \$2$ . The other numbers in Table 4 are derived in a similar way. The maximum expected payoff in each row coincides with the equilibrium bid, as indicated by an asterisk (\*). Note that the equilibrium involves bidding about one-half of the value.<sup>15</sup>

Table 5 shows the analogous calculations for the second treatment, with equally likely private values of \$0, \$3, or \$6. Interestingly, this increase in values does not alter the equilibrium bids in the unique Bayesian Nash equilibrium,

as indicated by the location of optimal bids for each value. Even though the equilibria are the same, we expected more of an upward bias in bids in the second (0, 3, 6) treatment. The intuition can be seen by looking at payoff losses associated with deviations from the Nash equilibrium. Consider, for instance, the middle-value bidder with expected payoffs shown in the second rows of Tables 4 and 5. In the (0, 3, 6) treatment, the cost of bidding \$1 above the equilibrium bid is  $\$1 - \$0.83 = \$0.17$ , which is less than the cost of bidding \$1 below the equilibrium bid:  $\$1 - \$0.50 = \$0.50$ . In the (0, 2, 5) treatment, the cost of an upward deviation from the equilibrium bid is greater than the cost of a downward deviation; see the middle row of Table 4. A similar argument applies to the high-value bidders, while deviation costs are the same in both treatments for the low-value bidder. Hence we expected more overbidding for the (0, 3, 6) treatment.

This intuition is borne out by bid data for the 50 subjects who participated in a single auction under each condition (again alternating the order of the two treatments and separating the two auctions with other one-shot games). Eighty percent of the bids in the (0, 2, 5) treatment matched the equilibrium: the average bids for low-, medium-, and high-value bidders were \$0, \$1.06, and \$2.64, respectively. In contrast, the average bids for the (0, 3, 6) treatment were \$0, \$1.82, and \$3.40 for the three value levels, and only 50 percent of all bids were Nash bids. The bid frequencies for each value are shown in Table 6. As in previous games, deviations from Nash behavior in these private-value auctions seem to be sensitive to the costs of deviation. Of

<sup>15</sup> The bids would be exactly one-half of the value if the highest value were \$4 instead of \$5, but we had to raise the highest value to eliminate multiple Nash equilibria.

TABLE 6—BID FREQUENCIES  
(EQUILIBRIUM BIDS MARKED WITH AN ASTERISK\*)

(0, 2, 5) Treatment			(0, 3, 6) Treatment		
	Bid	Frequency		Bid	Frequency
Value = 0	0*	20	Value = 0	0*	17
Value = 2	1*	15	Value = 3	1*	5
	2	1		2	11
	3	0		3	2
Value = 5	1	1	Value = 6	1	0
	2*	5		2*	3
	3	6		3	4
	4	2		4	6
	5	0		5	1
	6	0		6	1

course, this does not rule out the possibility that risk aversion or some other factor may also have some role in explaining the overbidding observed here, especially the slight overbidding for the high value in the (0, 2, 5) treatment.<sup>16</sup>

#### IV. Dynamic Games with Incomplete Information: Signaling

Signaling games are complex and interesting because the two-stage structure allows an opportunity for players to make inferences and change others' inferences about private information. This complexity often generates multiple equilibria that, in turn, have stimulated a sequence of increasingly complex refinements of the Nash equilibrium condition. Although it is unlikely that introspective thinking about the game will produce equilibrium behavior in a single play of a game this complex (except by coincidence), the one-shot play reveals useful information about subjects' cognitive processes.

In the experiment, half of the subjects were designated as "senders" and half as "responders." After reading the instructions, we began by throwing a die for each sender to determine

whether the sender was of type A or B. Everybody knew that the *ex ante* probability of a type A sender was one-half. The sender, knowing his/her own type would choose a signal, *Left* or *Right*. This signal determined whether the payoffs on the right or left side of Table 7 would be used. (The instructions used letters to identify the signals, but we will use words here to facilitate the explanations.) This signal would be communicated to the responder that was matched with that sender. The responder would see the sender's signal, *Left* or *Right*, but not the sender's type, and then choose a response, *C*, *D*, or *E*. The payoffs were determined by Table 7, with the sender's payoff to the left in each cell.

First, consider the problem facing a type A sender, for whom the possible payoffs from sending a *Left* signal (300, 0, 500) seem, in some loose sense, less attractive than those for sending a *Right* signal (450, 150, 1,000). For example, if each response is thought to be equally likely (the "principle of insufficient reason"), then the *Right* signal has a higher expected payoff. Consequently, type A's payoffs have been made bold for the *Right* row in the top right part of Table 7. Applying the principle of insufficient reason again, a type B sender looking at the payoffs in the bottom row of the table might be more attracted by the *Left* signal, with payoffs of (500, 300, 300) as compared with (450, 0, 0).<sup>17</sup> Therefore, sender B's payoffs are in bold for the *Left* signal. In fact, all of the type B subjects did send the *Left* signal, and seven of the ten type A subjects sent the *Right* signal. All responses in this game were *C*, so all but three of the outcomes were in one of the two cells marked by an asterisk. Notice that this is an equilibrium, since neither type of sender would benefit from sending the other signal, and the respondent cannot do any better than the maximum payoff received in the marked cells. This is a separating Nash equilibrium; the signal reveals the sender's type.

The payoff structure for this game becomes a little clearer if you think of the responses as one of three answers to a request: Concede, Deny, or

<sup>16</sup> Goeree et al. (2001) report a first-price auction experiment with six possible values, under repeated random matching for ten periods. A two-parameter econometric model that includes both decision error and risk aversion provides a good fit of 67 value/bid frequencies and shows that both the error parameter and risk-aversion parameter are significantly different from zero. David Lucking-Reiley (1999) mentions risk aversion as a possible explanation for overbidding in a variety of auction experiments.

<sup>17</sup> These are not dominance arguments, since the responder can respond differently to each signal, and the lowest payoff from sending one signal is not higher than the highest payoff that can be obtained from sending the other signal.

TABLE 7—SIGNALING WITH A SEPARATING EQUILIBRIUM (MARKED BY ASTERISKS) (SENDER'S PAYOFF, RESPONDER'S PAYOFF)

	Response to <i>Left</i> signal				Response to <i>Right</i> signal		
	<i>C</i>	<i>D</i>	<i>E</i>		<i>C</i>	<i>D</i>	<i>E</i>
Type A sends <i>Left</i>	300, 300	0, 0	500, 300	Type A sends <i>Right</i>	<b>450</b> , 900	<b>150</b> , 150	<b>1,000</b> , 300
Type B sends <i>Left</i>	<b>500</b> , 500	<b>300</b> , 450	<b>300</b> , 0	Type B sends <i>Right</i>	450, 0	0, 300	0, 150

TABLE 8—SIGNALING WITHOUT A SEPARATING EQUILIBRIUM (SENDER'S PAYOFF, RESPONDER'S PAYOFF)

	Response to <i>Left</i> signal				Response to <i>Right</i> signal		
	<i>C</i>	<i>D</i>	<i>E</i>		<i>C</i>	<i>D</i>	<i>E</i>
Type A sends <i>Left</i>	300, 300	0, 0	500, 300	Type A sends <i>Right</i>	<b>450</b> , 900	<b>150</b> , 150	<b>1,000</b> , 300
Type B sends <i>Left</i>	<b>300</b> , 300	<b>300</b> , 450	<b>300</b> , 0	Type B sends <i>Right</i>	450, 0	0, 300	0, 150

Evade. With some uncertainty about the sender's type, Evade is sufficiently unattractive to respondents that it is never selected. Consider the other two responses and note that a sender always prefers that the responder choose Concede instead of Deny. In the separating equilibrium, the signals reveal the senders' types, the responder always Concedes, and all players are satisfied. There is, however, a second equilibrium for the game in Table 7 in which the responder Concedes to *Left* and Denies *Right*, and therefore both sender types send *Left* to avoid being Denied.<sup>18</sup> Backward induction rationality (of the sequential Nash equilibrium) does not rule out these beliefs, since a deviation does not occur in equilibrium, and the responder is making a best response to the beliefs. What is unintuitive about these beliefs (that a deviant *Right* signal comes from a type B) is that the type B sender is earning 500 in this (*Left*, *Concede*) equilibrium outcome, and no deviation could conceivably increase this payoff. In contrast, the type A sender is earning 300 in the *Left* side pooling equilibrium, and this type could possibly earn more (450 or even 1,000), depending on the response to a deviation. The In-Koo Cho and Kreps (1987) *intui-*

*tive criterion* rules out these beliefs, and selects the separating equilibrium observed in the treasure treatment.<sup>19</sup>

The game in Table 8 is a minor variation on the previous game, with the only change being that the (500, 500) in the bottom left part of Table 7 is replaced by a (300, 300) payoff.<sup>20</sup> As before, consider the sender's expected payoffs when each response is presumed to be equally likely, which leads one to expect that type A senders will choose *Right* and that type B senders will choose *Left*, as indicated by the bold payoff numbers. In the experiment, 10 of the 13 type A senders did choose *Right*, and 9 of the 11 type B senders did choose *Left*. But the separation observed in this contradiction treatment is *not* a Nash equilibrium.<sup>21</sup> All equilibria for this

<sup>18</sup> To check that the responder has no incentive to deviate, note that Concede is a best response to a *Left* regardless of the sender's type, and that Deny is a best response to a deviant *Right* signal if the responder believes that it was sent by a type B.

<sup>19</sup> Brandts and Holt (1992, 1993) report experimental data that contradict the predictions of the intuitive criterion, i.e., the decision converged to an equilibrium ruled out by that criterion.

<sup>20</sup> Unlike the paired treatments considered previously, the payoff change for these signaling games does alter the set of Nash equilibria.

<sup>21</sup> The respondents would prefer to Concede to a *Right* signal and Deny a *Left* signal. Type B senders would therefore have an incentive to deviate from the proposed separating equilibrium and send a *Right* signal. In the experiment, half of the *Left* signals were Denied, whereas only 2 of the 12 *Right* signals were Denied.

contradiction treatment involve “pooling,” with both types sending the same signal.<sup>22</sup>

### V. Explaining Anomalous Behavior in One-Shot Games

Although the results for the contradiction treatments seem to preclude a game-theoretic explanation, many of the anomalous data patterns are related to the nature of the incentives. This suggests that it may be possible to develop formal models that explain both treasures and contradictions. Below we discuss several recent approaches that relax the common assumptions of perfect selfishness, perfect decision-making (no error), and perfect foresight (no surprises).

As noted in Section II, the anomalies observed for the dynamic games in Figures 3 and 4 are consistent with models of inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), which assumes that people like higher payoffs for themselves and dislike earning less than the other person (“envy”) or earning more (“guilt”). Inequity aversion also seems to play a role when players bargain over the division of a fixed amount of money (Goeree and Holt, 2000a). However, it cannot explain observed behavior in the contradiction matching pennies treatments. Consider, for example, the “320” version of the matching pennies game in Table 1. Since the column player is averse to the (320, 40) outcome, the column player would only be willing to randomize between *Left* and *Right* if the attractiveness of *Right* is increased by having the row player play *Bottom* more often than the 0.5 probability that would make a purely selfish column player indifferent. This prediction, that the row player should play *Bottom* more

often, is sharply contradicted by the data in the middle part of Table 1.<sup>23,24</sup>

Another possibility is that behavior in one-shot games conforms to a simple heuristic. Indeed, some experimental economists have suggested that subjects in the initial period of a repeated game choose the decision that maximizes their security level, i.e., the “maximum” decision. For example, in the Kreps game of Table 3, the frequently observed *Non-Nash* decision maximizes column’s security. The strong treatment effects in the matching pennies games cannot be explained in this way, however, since in all three treatments each player’s minimum payoff is the same for both decisions. A similar argument applies to the coordination game in Table 2. Moreover, the security-maximizing choices in the traveler’s dilemma and the minimum-effort coordination game are the lowest possible decision, which is contradicted by the high claim and effort choices in the contradiction treatments. Subjects may be risk averse in unfamiliar situations, but the extreme risk aversion implied by maximum security is generally not observed. Furthermore, heuristics based on reciprocity or a status quo bias do not apply to single-stage, one-shot games where there is neither a precedent nor an opportunity to reciprocate. Nor can loss aversion be the primary cause, since losses are impossible in most of the games reported here, and the possibility of a loss had no effect in the Kreps game.

As an alternative to simple heuristics, one could try to model players’ introspective thought processes. Previous models have typically specified some process of belief formation, assuming that players *best* respond to the

<sup>22</sup> For example, it is an equilibrium for both types to send *Right* if a *Left* signal triggers a *C* or a *D* response. The *D* response to *Left* is appropriate if the respondent thinks the deviant signal comes from a type B sender, and the *C* response is appropriate if the deviant is thought to be of type A. Beliefs that the deviant is of type A are intuitive, since type A earns 450 in equilibrium and could possibly earn more (500) by switching to *Left* (if an *E* response follows). A second pooling equilibrium involves both types sending a *Left* signal to which the respondents Concede. A deviant *Right* signal is Denied, which is appropriate if the respondent thinks the deviant signal comes from a type B sender. Again these beliefs are intuitive since the type B sender could possibly gain by deviating.

<sup>23</sup> Goeree et al. (2000) report formal econometric tests that reject the predictions of inequity aversion models in the context of a group of repeated asymmetric matching pennies games.

<sup>24</sup> Payoff inequity aversion also has no effect in the minimum-effort coordination game; any common effort level is still a Nash equilibrium. To see this, note that a unilateral effort increase from a common level reduces one’s own payoff and creates an disadvantageous inequity. Similarly, a unilateral decrease from a common effort level reduces one’s payoff and creates an inequity where one earns more than the other, since their costly extra effort is wasted. Thus inequity aversion cannot explain the strong effect of an increase in effort costs.



resulting beliefs.<sup>25</sup> The experiments reported above indicate that *magnitudes* (not just *signs*) of payoff differences matter, and it is thus natural to consider a decision rule for which choice probabilities are positively but imperfectly related to payoffs. The logit rule, for example, specifies that choice probabilities,  $p_i$ , for options  $i = 1, \dots, m$ , are proportional to exponential functions of the associated expected payoffs,  $\pi_i^e$ :

$$(1) \quad p_i = \frac{\exp(\pi_i^e/\mu)}{\sum_{j=1, \dots, m} \exp(\pi_j^e/\mu)}, \quad i = 1, \dots, m,$$

where the sum in the denominator ensures that the probabilities sum to one, and the “error parameter,”  $\mu$ , determines how sensitive choice probabilities are to payoff differences.<sup>26</sup>

In order to use the “logit best response” in (1), we need to model the process of belief formation, since belief probabilities are used to calculate the expected payoffs on the right side of (1). By the principle of insufficient reason one might postulate that each of the others’ actions are equally likely. This corresponds to the Stahl and Wilson (1995) notion of “level one” rationality, which captures many of the first-period decisions in the “guessing game”

<sup>25</sup> Perhaps the best-known model of introspection is Harsanyi and Selten’s (1988) “tracing procedure.” This procedure involves an axiomatic determination of players’ common priors (the “preliminary theory”) and the construction of a modified game with payoffs for each decision that are weighted averages of those in the original game and of the expected payoffs determined by the prior distribution. By varying the weight on the original game, a sequence of best responses for the modified game are generated. This process is used to select one of the Nash equilibria of the original game. Gonzalo Olcina and Amparo Urbano (1994) also use an axiomatic approach to select a prior distribution, which is then revised by a simulated learning process that is essentially a partial adjustment from current beliefs to best responses to current beliefs. Since neither the Harsanyi/Selten model nor the Olcina/Urbano model incorporates any noise, they predict that behavior will converge to the Nash equilibrium in games with a unique equilibrium, which is an undesirable feature in light of the contradictions data reported above.

<sup>26</sup> As  $\mu$  goes to zero, payoff differences are “blown up,” and the probability of the optimal decision converges to 1. In the other extreme, as  $\mu$  goes to  $\infty$ , the choice probabilities converge to  $1/m$  independently of expected payoffs. See R. Duncan Luce (1959) for an axiomatic derivation of the logit choice rule in (1).

reported by Nagel (1995).<sup>27</sup> It is easy to verify that level one rationality also provides good predictions for both treasure and contradiction treatments in the traveler’s dilemma, the minimum-effort coordination game, and the Kreps game. There is evidence, however, that at least some subjects form more precise beliefs about others’ actions, possibly through higher levels of introspection.<sup>28</sup> In the matching pennies games in Table 1, for example, a flat prior makes column indifferent between *Left* and *Right*, and yet most column players seem to anticipate that row will choose *Top* in the 320 version and *Bottom* in the 44 version of this game.

Of course, what the other player does depends on what they think you will do, so the next logical step is to assume that *others* make responses to a flat prior, and then you respond to that anticipated response (Selten, 1991). This is Stahl and Wilson’s (1995) “level two” rationality. There is, however, no obvious reason to truncate the levels of iterated thinking. The notion of rationalizability discussed above, for example, involves infinitely many levels of iterated thinking, with “never-best” responses eliminated in succession. But rationalizability seems to imply too much rationality, since it predicts that all claims in the traveler’s dilemma will be equal to the minimum claim, independent of the penalty/reward parameter. One way to limit the precision of the thought process, without making an arbitrary assumption about the number of iterations, is to inject increasing amounts of noise into higher levels of iterated thinking (Goeree and Holt, 1999; Kübler and Weizsäcker, 2000). Let  $\phi_\mu$  denote the logit best-response map (for error rate  $\mu$ ) on the right side of (1). Just as a single logit response to beliefs,  $\mathbf{p}_0$ , can be represented as  $p = \phi_\mu(\mathbf{p}_0)$ , a series of such responses can be represented as:<sup>29</sup>

<sup>27</sup> In our own work, we have used a noisy response to a flat prior as a way of starting computer simulations of simulations of behavior in repeated games (Brandts and Holt, 1996; Capra et al., 1999, 2002; Goeree and Holt, 1999).

<sup>28</sup> Costa-Gomes et al. (2001), for example, infer some heterogeneity in the amount of introspection by observing the types of information that subjects acquire before making a decision.

<sup>29</sup> Goeree and Holt (2000b) use continuity arguments to show that the limit in (2) exists even if the (increasing) error parameters are person specific.

$$(2) \quad p = \lim_{n \rightarrow \infty} \phi_{\mu_1}(\phi_{\mu_2}(\dots \phi_{\mu_n}(p_0))),$$

where  $\mu_1 \leq \mu_2 \leq \dots$ , with  $\mu_\infty$  converging to infinity.<sup>30</sup> This assumption captures the idea that it becomes increasingly complex to do more and more iterations.<sup>31</sup> Since  $\phi_\mu$  for  $\mu = \infty$  maps the whole probability simplex to a single point, the right side of (2) is independent of the initial belief vector  $p_0$ . Moreover, the introspection process in (2) yields a unique outcome even in games with multiple Nash equilibria. Note that the choice probabilities on the left side of (2) generally do not match the beliefs at any stage of the iterative process on the right. In other words, the introspective process allows for surprises, which are likely to occur in one-shot games.

For games with very different levels of complexity such as the ones reported here, the error parameters that provide the best fit are likely to be different. In this case, the estimates indicate the degree of complexity, i.e., they serve as a measurement device. For games of similar complexity, the model in (2) could be applied to predict behavior across games. We have used it to explain data patterns in a series of 37 simple matrix games, assuming a simple two-parameter model for which  $\mu_n = \mu t^n$ , where  $t$  determines the rate at which noise increases with higher iterations (Goeree and Holt, 2000b). The estimated value ( $t = 4.1$ ) implies that there is more noise for higher levels of introspection, a result that is roughly consistent with estimates obtained by Kübler and Weizsäcker (2000) for data from information-cascade experiments.

The analysis of introspection is a relatively understudied topic in game theory, as compared with equilibrium refinements and learning, for

example. Several of the models discussed above do a fairly good job of organizing the qualitative patterns of conformity and deviation from the predictions of standard theory, but there are obvious discrepancies. We hope that this paper will stimulate further theoretical work on models of behavior in one-shot games. One potentially useful approach may be to elicit beliefs directly as the games are played (Theo Offerman, 1997; Andrew Schotter and Yaw Narkov, 1998).

## VI. Conclusion

One-shot game experiments are interesting because many games are in fact only played once; single play is especially relevant in applications of game theory in other fields, e.g., international conflicts, election campaigns, and legal disputes. The decision makers in these contexts, like the subjects in our experiments, typically have experience in similar games with other people. One-shot games are also appealing because they allow us to abstract away from issues of learning and attempts to manipulate others' beliefs, behavior, or preferences (e.g., reciprocity, cooperativeness). This paper reports the results of ten pairs of games that are played only once by subjects who have experience with other one-shot and repeated games. The Nash equilibrium (or relevant refinement) provides accurate predictions for standard versions of these games. In each case, however, there is a matched game for which the Nash prediction clearly fails, although it fails in a way that is consistent with simple (non-game-theoretic) intuition. The results for these experienced subjects show:

- (1) Behavior may diverge sharply from the unique rationalizable (Nash) equilibrium in a social (traveler's) dilemma. In these games, the Nash equilibrium is located on one side of the range of feasible decisions, and data for the contradiction treatment have a mode on the *opposite* side of this range. The most salient feature of the data is the extreme sensitivity to a parameter that has no effect on the Nash outcome.
- (2) Students suffering through game theory classes may have good reasons when they have trouble understanding why a change in one player's payoffs only affects the

<sup>30</sup> The case of a constant parameter ( $\mu_1 = \mu_2 = \dots = \mu$ ) is of special interest. In this case, the process may not converge for some games (e.g., matching pennies), but when it does, the limit probabilities,  $p^*$ , must be invariant under the logit map:  $\phi_\mu(p^*) = p^*$ . A fixed point of this type constitutes a "logit equilibrium," which is a special case of the quantal-response equilibrium defined in McKelvey and Palfrey (1995). It is in this sense that the logit equilibrium arises as a limit case of the noisy introspective process defined in (2).

<sup>31</sup> For an interesting alternative approach, see Capra (1998). In her model, beliefs are represented by degenerate distributions that put all probability mass at a single point. The location of the belief points is, *ex ante*, stochastic.

other player's decision probabilities in a mixed-strategy Nash equilibrium. The data from matching pennies experiments show strong "own-payoff" effects that are not predicted by the unique (mixed-strategy) Nash equilibrium. The Nash analysis seems to work only by coincidence, when the payoff structure is symmetric and deviation risks are balanced.

- (3) Effort choices are strongly influenced by the cost of effort in coordination games, an intuitive result that is not explained by standard theory, since *any* common effort is a Nash equilibrium in such games. Moreover, as Kreps conjectured, it is possible to design coordination games where the majority of one player's decisions correspond to the only action that is not part of any Nash equilibrium.
- (4) Subjects often do not trust others to be rational when irrationality is relatively costless. Moreover, "threats" that are *not* credible in a technical sense may nevertheless alter behavior in simple two-stage games when carrying out these threats is not costly.
- (5) Deviations from Nash predictions in alternating-offer bargaining games and in private-value auctions are inversely related to the costs of such deviations. The effects of these biases can be quite large in the games considered.
- (6) It is possible to set up a simple signaling game in which the decisions reveal the signaler's type (separation), even though the equilibrium involves pooling.

So what should be done? Reinhard Selten, one of the three game theorists to share the 1994 Nobel Prize, has said: "Game theory is for proving theorems, not for playing games."<sup>32</sup> Indeed, the internal elegance of traditional game theory is appealing, and it has been defended as being a normative theory about how perfectly rational people should play games with each other, rather than a positive theory that predicts actual behavior (Rubinstein, 1982). It is natural to separate normative and positive studies of individual decision-making, which allows one to

compare actual and optimal decision-making. This normative-based defense is not convincing for games, however, since the best way for one to play a game depends on how others *actually* play, not on how some theory dictates that rational people *should* play. John Nash, one of the other Nobel recipients, saw no way around this dilemma, and when his experiments were not providing support to theory, he lost whatever confidence he had in the relevance of game theory and focused on more purely mathematical topics in his later research (Nasar, 1998).

Nash seems to have undersold the importance of his insight, and we will be the first to admit that we begin the analysis of a new strategic problem by considering the equilibria derived from standard game theory, before considering the effects of payoff and risk asymmetries on incentives to deviate. But in an interactive, strategic context, biases can have reinforcing effects that drive behavior well away from Nash predictions, and economists are starting to explain such deviations using computer simulations and theoretical analyses of learning and decision error processes. There has been relatively little theoretical analysis of one-shot games where learning is impossible. The models of iterated introspection discussed here offer some promise in explaining the qualitative features of deviations from Nash predictions enumerated above. Taken together, these new approaches to a stochastic game theory enhance the behavioral relevance of standard game theory. And looking at laboratory data is a lot less stressful than before.

## REFERENCES

- Anderson, Simon P.; Goeree, Jacob K. and Holt, Charles A. "Rent Seeking with Bounded Rationality: An Analysis of the All-Pay Auction." *Journal of Political Economy*, August 1998a, 106(4), pp. 828–53.
- \_\_\_\_\_. "A Theoretical Analysis of Altruism and Decision Error in Public Goods Games." *Journal of Public Economics*, November, 1998b, 70(2), pp. 297–323.
- \_\_\_\_\_. "The Logit Equilibrium: A Perspective on Intuitive Behavioral Anomalies." *Southern Economic Journal*, 2001a (forthcoming).
- \_\_\_\_\_. "Minimum-Effort Coordination

<sup>32</sup> Selten reiterated this point of view in a personal communication to the authors.

- Games: Stochastic Potential and Logit Equilibrium." *Games and Economic Behavior*, February 2001b, 34(2), pp. 177–99.
- Basu, Kaushik.** "The Traveler's Dilemma: Paradoxes of Rationality in Game Theory." *American Economic Review*, May 1994 (*Papers and Proceedings*), 84(2), pp. 391–95.
- Beard, Randolph T. and Beil, Richard O.** "Do People Rely on the Self-interested Maximization of Others: An Experimental Test." *Management Science*, February 1994, 40(2), pp. 252–62.
- Bernheim, B. Douglas.** "Rationalizable Strategic Behavior." *Econometrica*, July 1984, 52(4), pp. 1007–28.
- Bolton, Gary E.** "Bargaining and Dilemma Games: From Laboratory Data Towards Theoretical Synthesis." *Experimental Economics*, 1998, 1(3), pp. 257–81.
- Bolton, Gary E and Ockenfels, Axel.** "A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, March 2000, 90(1), pp. 166–93.
- Brandts, Jordi and Holt, Charles A.** "An Experimental Test of Equilibrium Dominance in Signaling Games." *American Economic Review*, December 1992, 82(5), pp. 1350–65.
- \_\_\_\_\_. "Adjustment Patterns and Equilibrium Selection in Experimental Signaling Games." *International Journal of Game Theory*, 1993, 22(3), pp. 279–302.
- \_\_\_\_\_. "Naive Bayesian Learning and Adjustments to Equilibrium in Signaling Games." Working paper, University of Virginia, 1996.
- Camerer, Colin and Ho, Teck-Hua.** "Experience-Weighted Attraction Learning in Normal Form Games." *Econometrica*, July 1999, 67(4), pp. 827–74.
- Capra, C. Monica.** "Noisy Expectation Formation in One-Shot Games." Ph.D. dissertation, University of Virginia, 1998.
- Capra, C. Monica; Goeree, Jacob K.; Gomez, Rosario and Holt, Charles A.** "Anomalous Behavior in a Traveler's Dilemma?" *American Economic Review*, June 1999, 89(3), pp. 678–90.
- \_\_\_\_\_. "Learning and Noisy Equilibrium Behavior in an Experimental Study of Imperfect Price Competition." *International Economic Review*, 2002 (forthcoming).
- Cho, In-Koo and Kreps, David M.** "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics*, May 1987, 102(2), pp. 179–221.
- Cooper, David J.; Garvin, Susan and Kagel, John H.** "Signaling and Adaptive Learning in an Entry Limit Pricing Game." *Rand Journal of Economics*, Winter 1997, 28(4), pp. 662–83.
- Cooper, Russell; DeJong, Douglas V.; Forsythe, Robert and Ross, Thomas W.** "Communication in Coordination Games." *Quarterly Journal of Economics*, May 1992, 107(2), pp. 739–71.
- Costa-Gomes, Miguel; Crawford, Vincent P. and Broseta, Bruno.** "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica*, September 2001, 69, pp. 1193–235.
- Costa-Gomes, Miguel and Zauner, Klaus G.** "Ultimatum Bargaining Behavior in Israel, Japan, Slovenia, and the United States: A Social Utility Analysis." *Games and Economic Behavior*, February 2001, 34(2), pp. 238–69.
- Crawford, Vincent P.** "Adaptive Dynamics in Coordination Games." *Econometrica*, January 1995, 63(1), pp. 103–43.
- Davis, Douglas D. and Holt, Charles A.** *Experimental economics*. Princeton, NJ: Princeton University Press, 1993.
- Eckel, Catherine and Wilson, Rick.** "The Human Face of Game Theory: Trust and Reciprocity in Sequential Games." Working paper, Rice University, 1999.
- Erev, Ido and Roth, Alvin E.** "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria." *American Economic Review*, September 1998, 88(4), pp. 848–81.
- Fehr, Ernst and Schmidt, Klaus.** "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, August 1999, 114(3), pp. 817–68.
- Fudenberg, Drew and Levine, David K.** *The theory of learning in games*. Cambridge, MA: MIT Press, 1998.
- Fudenberg, Drew and Tirole, Jean.** *Game theory*. Cambridge, MA: MIT Press, 1993.
- Gibbons, Robert.** "An Introduction to Applicable Game Theory." *Journal of Economic Perspectives*, Winter 1997, 11(1), pp. 127–49.
- Goeree, Jacob K. and Holt, Charles A.** "An Experimental Study of Costly Coordination." Working paper, University of Virginia, 1998.

- \_\_\_\_\_. "Stochastic Game Theory: For Playing Games, Not Just for Doing Theory." *Proceedings of the National Academy of Sciences*, September 1999, 96, pp. 10,564–67.
- \_\_\_\_\_. "Asymmetric Inequality Aversion and Noisy Behavior in Alternating-Offer Bargaining Games." *European Economic Review*, May 2000a, 44(4–6), pp. 1079–89.
- \_\_\_\_\_. "Models of Noisy Introspection." Working paper, University of Virginia, 2000b.
- Goeree, Jacob K.; Holt, Charles A. and Palfrey, Thomas R.** "Risk-Averse Behavior in Asymmetric Matching Pennies Games." Working paper, University of Virginia, 2000.
- \_\_\_\_\_. "Quantal Response Equilibrium and Overbidding in Private Value Auctions." *Journal of Economic Theory*, 2001 (forthcoming).
- Guarnaschelli, Serena; McKelvey, Richard D. and Palfrey, Thomas R.** "An Experimental Study of Jury Decision Making." *American Political Science Review*, 2001 (forthcoming).
- Harrison, Glenn W.** "Theory and Misbehavior of First-Price Auctions." *American Economic Review*, September 1989, 79(4), pp. 749–62.
- Harsanyi, John C. and Selten, Reinhard.** *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press, 1988.
- Kahneman, Daniel; Knetsch, Jack L. and Thaler, Richard H.** "The Endowment Effect, Loss Aversion, and Status Quo Bias: Anomalies." *Journal of Economic Perspectives*. Winter 1991, 5(1), pp. 193–206.
- Kahneman, Daniel; Slovic, Paul and Tversky, Amos.** *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press, 1982.
- Kreps, David M.** "Nash Equilibrium," in John Eatwell, Murray Milgate, and Peter Newman, eds., *The new Palgrave: Game theory*. New York: Norton, 1995, pp. 176–77.
- Kübler, Dorothea and Weizsäcker, Georg.** "Limited Depth of Reasoning and Failure of Cascade Formation in the Laboratory." Working paper, Harvard University, 2000.
- Luce, R. Duncan.** *Individual choice behavior*. New York: Wiley, 1959.
- Lucking-Reiley, David.** "Using Field Experiments to Test the Equivalence Between Auction Formats: Magic on the Internet." *American Economic Review*, December 1999, 89(5), pp. 1063–80.
- Mailath, George.** "Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory." *Journal of Economic Literature*. September 1998, 36(3), pp. 1347–74.
- McKelvey, Richard D. and Palfrey, Thomas R.** "An Experimental Study of the Centipede Game." *Econometrica*, July 1992, 60(4), pp. 803–36.
- \_\_\_\_\_. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior*. July 1995, 10(1), pp. 6–38.
- \_\_\_\_\_. "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics*, 1998, 1(1), pp. 9–41.
- McKelvey, Richard D.; Palfrey, Thomas R. and Weber, Roberto A.** "The Effects of Payoff Magnitude and Heterogeneity on Behavior in  $2 \times 2$  Games with Unique Mixed Strategy Equilibria." *Journal of Economic Behavior and Organization*. August 2000, 42(4), pp. 523–48.
- Mookherjee, Dilip and Sopher, Barry.** "Learning and Decision Costs in Experimental Constant Sum Games." *Games and Economic Behavior*, April 1997, 19(1), pp. 97–132.
- Nagel, Rosemarie.** "Unraveling in Guessing Games: An Experimental Study." *American Economic Review*, December 1995, 85(5), pp. 1313–26.
- Nasar, Sylvia.** *A beautiful mind*. New York: Simon and Schuster, 1998.
- Nash, John.** "The Bargaining Problem." *Econometrica*, April 1950, 18(2), pp. 155–62.
- Ochs, Jack.** "Games with Unique Mixed Strategy Equilibria: An Experimental Study." *Games and Economic Behavior*, July 1995a, 10(1), pp. 202–17.
- \_\_\_\_\_. "Coordination Problems," in John H. Kagel and Alvin E. Roth, eds., *Handbook of experimental economics*. Princeton, NJ: Princeton University Press, 1995b, pp. 195–249.
- Offerman, Theo.** *Beliefs and decision rules in public goods games: Theory and experiments*. Dordrecht: Kluwer Academic Press, 1997.
- Olcina, Gonzalo and Urbano, Amparo.** "Introspection and Equilibrium Selection in  $2 \times 2$  Matrix Games." *International Journal of Game Theory*, 1994, 23(3), pp. 183–206.
- Pearce, David G.** "Rationalizable Strategic

- Behavior and the Problem of Perfection." *Econometrica*, July 1984, 52(4), pp. 1029–50.
- Rabin, Matthew.** "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, December 1993, 83(5), pp. 1281–302.
- Reynolds, Stanley S.** "Sequential Bargaining with Asymmetric Information: The Role of Quantal Response Equilibrium in Explaining Experimental Results." Working paper, University of Arizona, 1999.
- Rosenthal, Robert W.** "Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox." *Journal of Economic Theory*, August 1981, 25(1), pp. 92–100.
- Roth, Alvin E.** "Bargaining Experiments," in John H. Kagel and Alvin E. Roth, eds., *Handbook of experimental economics*. Princeton, NJ: Princeton University Press, 1995, pp. 253–348.
- Rubinstein, Ariel.** "Perfect Equilibrium in a Bargaining Model." *Econometrica*, January 1982, 50(1), pp. 97–109.
- Schotter, Andrew and Narkov, Yaw.** "Equilibria in Beliefs and Our Belief in Equilibrium: An Experimental Approach." Working paper, New York University, 1998.
- Selten, Reinhard.** "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory*, 1975, 4, pp. 25–55.
- \_\_\_\_\_. "Anticipatory Learning in Two-Person Games," in Reinhard Selten, ed., *Game equilibrium models I*. Berlin: Springer-Verlag, 1991, pp. 98–154.
- Selten, Reinhard and Buchta, Joachim.** "Experimental Sealed Bid First Price Auctions with Directly Observed Bid Functions" in I. E. D. Budescu and R. Zwick, eds., *Games and human behavior: Essays in honor of Amnon Rapoport*. Hillside, NJ: Erlbaum Association, 1998.
- Stahl, Dale O. and Wilson, Paul W.** "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, July 1995, 10(1), pp. 218–54.
- Straub, Paul G.** "Risk Dominance and Coordination Failures in Static Games." *Quarterly Review of Economics and Finance*. Winter 1995, 35(4), pp. 339–63.
- Tucker, Al W.** "A Two-Person Dilemma." Mimeo, Stanford University, 1950. [Published under the heading "On Jargon: The Prisoner's Dilemma." *UMAP Journal*, 1980, 1, p. 101.]
- Van Huyck, John B.; Battalio, Raymond C. and Beil, Richard O.** "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure." *American Economic Review*, March 1990, 80(1), pp. 234–48.
- Vickrey, William.** "Counterspeculation, Auctions, and Sealed Tenders." *Journal of Finance*, 1961, 16, pp. 8–37.
- von Neumann, John and Morgenstern, Oscar.** *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press, 1944.