American Economic Review 2009, 99:4, 1619–1635
http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.4.1619

Field Centipedes

By Ignacio Palacios-Huerta and Oscar Volij*

The centipede game is perhaps the best example of what is known as "paradoxes of backward induction." These paradoxes involve sequential games, all of whose correlated equilibria, and *a fortiori* all its Nash equilibria, imply a very counterintuitive play.

A particular instance of the centipede game can be described as follows. A pile with four dollar bills and another with one dollar bill are lying on a table. Player 1 has two options, either to "stop" or to "continue." If he stops, the game ends and he gets \$4 while Player 2 gets the remaining dollar. If he continues, the two piles are doubled to \$8 and \$2, and Player 2 is faced with a similar decision: either to take the larger pile (\$8), thus ending the game and leaving the smaller pile (\$2) for Player 1, or to let the piles double again and let Player 1 decide. The game continues for at most six rounds. If, by then, neither of the players has stopped, Player 1 gets \$256 and Player 2 gets \$64. Figure 1 depicts this situation.

Although this game offers both players a very profitable opportunity, all standard game theoretic solution concepts predict that Player 1 will stop at the first opportunity, and win just \$4. Despite this unambiguous prediction, game theorists often "wonder if it really reflects the way in which *anyone* would play such a game" (Richard D. McKelvey and Thomas R. Palfrey 1992, 804, italics added).

The game theoretic prescription for this kind of sequential games goes so much against intuition that it induced Robert W. Rosenthal (1981), in the same paper in which he introduced the centipede game, to propose an alternative to the game theoretic approach in the hope of obtaining predictions more in line with intuition.[1] Robert J. Aumann (1992) contends that the backward induction outcome in these games is so disturbing to some people that "if this is rationality, they want none of it" (218).

The apparent conflict between the theoretical prediction and intuitively reasonable behavior in the centipede game prompted some researchers to argue that there may not be any conflict between rationality and the failure of backward induction. In a very convincing example, Aumann (1992) shows that in the centipede game it is possible for rationality to be mutually known to a high degree (in fact, the rationality of one of the players may even be commonly known) and still for both players to "continue" for several rounds. Phil J. Reny (1992) also eloquently demonstrates how violating backward induction may be perfectly rational. Elhanan Ben-Porath (1997) shows that several rounds of "continuation" are consistent with common certainty of rationality.[2] Therefore, rationality alone does not imply the pessimistic and rather unprofitable behavior prescribed by the game theoretic solution concepts.

* Palacios-Huerta: Department of Management, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom (e-mail: i.palacios-huerta@lse.ac.uk); Volij: Department of Economics, Ben-Gurion University, Beer-Sheva 84105, Israel (e-mail: ovolij@bgu.ac.il). We thank three anonymous referees, Jose Apesteguia, Pedro Dal Bó, Martin Dufwenberg, Leontxo García, Julio Gonzalez-Díaz, Yona Rubinstein, Ana Saracho, and participants in various seminars and conferences for helpful comments. We are especially indebted to the organizers of the XXII Open International Chess Tournament of Sestao, the X Open International Chess Tournament of León, and the XXVI Open International Chess Tournament Villa de Benasque for access to the participants in their tournaments. We gratefully acknowledge financial support from the Salomon Foundation and the Spanish Ministerio de Ciencia y Tecnología (grants BEC2003-08182 and SEJ2006-05455), as well as the editing assistance from Estelle Shulgasser.

[1] While Rosenthal's proposal did not catch on in the literature, his centipede game has become a cornerstone example of the conflict between theory and intuition.

[2] See also Geir B. Asheim and Martin Dufwenberg (2003) for a refinement of this result.

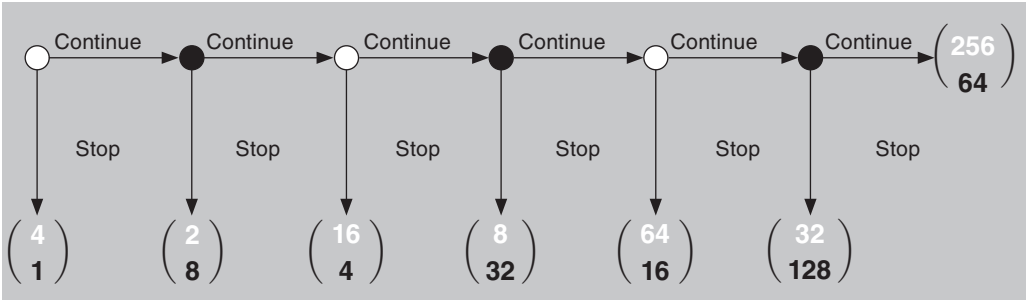

Figure 1. A Centipede Game

It actually turns out that it is not rationality, nor even mutual knowledge of rationality, but *common knowledge of rationality* that implies the backward induction outcome. Indeed, Aumann (1995) formalizes a notion of rationality in perfect information games that allows him to make this statement precise.[3] However, he also concedes that common knowledge of rationality "is an ideal condition that is rarely met in practice" (18), and further contends that if this condition is absent, the backward induction outcome need not emerge. In particular, he stresses that in the centipede game even the smallest departure from common knowledge of rationality may induce rational players to depart significantly from equilibrium play.

In the next section we review the empirical evidence in this game. Consistent with intuition, a number of experimental studies conducted with college students have documented systematic departures from the backward induction outcome, typically finding that almost no subjects stop at the first opportunity, even after they have played several repetitions of the game. Further, these studies often conjecture that various forms of social preferences, limited cognition, or failures of backward induction reasoning play an important role in explaining why the equilibrium outcome is rarely observed in the lab.

In this paper we depart from previous experimental studies in the subject pool we consider. We first identify subjects who are very likely characterized by a high degree of rationality, namely expert chess players. These players devote a large part of their life to finding optimal strategies for innumerable chess positions using backward induction reasoning. More important, one can safely say that it is common knowledge among most humans that chess players are highly familiar with backward induction reasoning. Our purpose is to use these subjects to study the extent to which knowledge of an opponent's rationality is a key determinant of the predictive power of subgame-perfect equilibrium in this game. By varying the "closeness" to common knowledge of rationality across different experimental treatments, we design a test that can separate the hypothesis of the epistemic literature on rationality from that of social preferences. More precisely, social preferences would imply that the results are roughly the same across different treatments, while the epistemic approach would suggest the results to be closer to equilibrium the "closer" we are to common knowledge of rationality. We investigate this question both in a field and in a lab experiment.

Our first experiment takes place in the field, where chess players were matched with each other at various chess tournaments. Each chess player participated in the experiment only once, playing only one round of the centipede game. Our second and main experiment takes place in a lab

[3] Using a different formalization, Reny (1993) shows that the backward induction outcome may fail to occur even if there is common knowledge of rationality at the beginning of the game. See also Ben-Porath (1997) and Asheim and Dufwenberg (2003).

setting where both chess players and students were matched with either chess players or students in four treatments. These treatments differ in the order of play of these two types of subjects. In this experiment subjects play ten rounds of centipede game, and no subject plays against the same opponent twice.

Our main findings are the following:

(i) Both in the field and in the lab, when chess players play against chess players, the outcome is very close to the subgame-perfect equilibrium prediction. In the field experiment with chess players playing a one-shot centipede game, 69 percent of the games ended at the initial node. When we restrict attention to games where the first player was a Grandmaster, this percentage escalates to 100 percent. In the laboratory experiment, when chess players play ten repetitions of the centipede game against chess players, more than 70 percent of the games ended at the first node. More importantly, we find that *every* chess player converges fully to equilibrium play already at the fifth repetition. These results suggest that the ideal condition of common knowledge of rationality seems to be approached closely when chess players play the centipede game.

(ii) When students play against chess players in our laboratory experiment, the outcome is much closer to the subgame-perfect equilibrium than when students play against students. More precisely, when students played against students their behavior was consistent with previous experimental results. Only 3 percent of the games ended at the initial node, and there was no sign of convergence to equilibrium play as the repetitions progressed. In the treatments where students faced chess players and acted as Player 1, the proportion of games that ended at the first node increased tenfold, to 30 percent. Furthermore, when we restrict attention to the last two repetitions, this proportion grows to 70 percent. Lastly, when chess players acted as Player 1 and students acted as Player 2, 37.5 percent of the games ended in the first node.

We view these findings as being highly consistent with the predictions of the theoretical literature in that the predictive power of subgame-perfect equilibrium hinges mainly on knowledge of players' rationality, and not on altruism or social preferences. Hence, the results offer strong support for standard approaches to economic modeling based on the principles of self-interested rational economic agents and on their assessments of the behavior of their opponents in a game. Thus, at a moment when there is much discussion about nonstandard assumptions on players' preferences, the results in this paper suggest that such assumptions might be neither a realistic nor even a necessary modeling device.

The rest of the paper is organized as follows. Section I briefly reviews the experimental literature on the centipede game and backward induction. Sections II and III describe our field and laboratory experiments, respectively, and their results. Section IV concludes.

I. Literature Review

Uneasiness with the backward induction outcome arose long before the first experimental study of the centipede game was performed. Indeed, McKelvey and Palfrey (1992) begin their pioneering paper by stating that they report on experimental games whose Nash equilibrium predictions "are widely acknowledged to be unsatisfactory." These experiments resulted in outcomes so distant from the game theoretic predictions that the intuition against the backward induction outcome seemed to be conclusively vindicated: fewer than 1.5 percent of the games played in McKelvey-Palfrey's centipede game experiment resulted in the backward induction

outcome, even after subjects played several repetitions of the game, and these findings have been confirmed in other studies.[4]

There were some later attempts to experimentally test the backward induction prediction in centipede-like games. One is based on the idea that since the pie to be divided between the players in these games grows as play advances to later nodes, the tendency not to exit at early nodes could be explained by means of a small measure of altruism. Mark Fey, McKelvey, and Palfrey (1996) ran a series of experiments with *constant-sum* centipede games. These are games where the amount to be divided is constant, and only its distribution among the players becomes more and more unequal as play moves forward. As in the regular centipede game, this constant-sum game has a unique Nash equilibrium outcome, which results in an immediate "stop." Since moderate altruism cannot induce players to "continue" at their respective decision nodes, one would expect a high proportion of these games to result in the backward induction outcome.[5] Indeed, when two kinds of constant-sum centipede games were run, one with ten nodes and a second with six nodes, the proportion of games that resulted in the backward induction outcome was 45 percent and 59 percent, respectively. Although this is a dramatic increase in the performance of the theoretical prediction, Fey, McKelvey, and Palfrey (1996) still regarded backward induction as inadequate to explain players' behavior.

Another attempt at achieving the backward induction outcome was more recently implemented by Amnon Rapoport et al. (2003). They ran a series of three-person centipede games with substantially higher payoffs and many more repetitions than in the original McKelvey and Palfrey (1992) experiment (60 rounds rather than 10). Here, again, the backward induction outcome was observed to be played more often (46 percent of the trials) than in McKelvey-Palfrey's experiment, but nonetheless was not enough to support the theoretical predictions. Interestingly, in the last five repetitions of this 60-round, three-player, high-stakes experiment, 75 percent of the games ended in the initial node. This seems to be consistent with the idea that substantial experience from repeated play in stable settings, especially in high-stakes situations, may lead to the backward induction outcome.

The experiments we implement in this paper are quite different from the ones described above. Our experiment in the field represents a novel strategic situation in which subjects play only once. Thus, the design suppresses learning and repeated-game effects, and elicits subjects' "initial responses." As in Miguel A. Costa-Gomes and Vincent P. Crawford (2006), this allows us to study strategic thinking "uncontaminated" by learning. On the other hand, in our laboratory, experiment learning is not the main focus of the analysis and hence we allow a standard small number of repetitions. As indicated earlier, by simply introducing subjects who are likely characterized by a high degree of rationality into an otherwise standard design with college students, our purpose is to study the extent to which knowledge of opponent's rationality is a key determinant of the predictive power of subgame-perfect equilibrium. Put differently, we are interested in the comparative statics suggested by the epistemic approach.

Finally, the failure of the equilibrium model to predict the outcomes of past experiments prompted researchers to propose and test alternative models of strategic behavior in this game. Two of these approaches, both of which involve the introduction of a slight perturbation to the original game, are the following. The first one transforms the original centipede game into a

[4] For instance, Rosemarie Nagel and Fang Fang Tang (1998) implement an experiment on the centipede game played in reduced normal form. Even after subjects repeat the game 100 times against randomly selected opponents, fewer than 1 percent of the games end in the backward induction outcome. Gary Bornstein, Tamar Kugler, and Anthony Ziegelmeyer (2004) find in their sample that even if individuals play in groups, no games end in any of the first two nodes.

[5] By moderate altruism, we mean other-regarding preferences according to which a dollar to oneself is preferable to the same dollar belonging to the other.

game of imperfect information by introducing an altruistic type with small but positive probability, and then calculates its sequential equilibrium.[6] In this modified game, each player assigns a positive probability that his opponent is an altruistic type who continues at every node. The resulting game has a unique sequential equilibrium, which depends on the prior probability of the altruistic type. This equilibrium requires the nonaltruistic players to continue with positive probability at every node, except for the last one. The reason for this behavior is that the mere possibility of the existence of altruistic players allows the nonaltruistic players to mimic the altruistic behavior.[7] McKelvey and Palfrey (1992) used a version of this model to account for most features of their experimental data.

The second approach is based on the quantal response equilibrium concept introduced by McKelvey and Palfrey (1995) for the analysis of normal form games, or rather its version for extensive form games of perfect information known as the *agent quantal response equilibrium* (AQRE) introduced by McKelvey and Palfrey (1998). The AQRE model is a generalization of the standard equilibrium model in which agents evaluate the payoffs of each possible strategy combination according to random perturbations of the original payoffs. Specifically, the AQRE is a Nash equilibrium of the perturbation of the original game and coincides with it when the perturbation vanishes.

McKelvey and Palfrey (1998) use this model for the analysis of their 1992 experimental data. A very similar specification is used by Klaus G. Zauner (1999). Fey, McKelvey, and Palfrey (1996) use the AQRE model in their analysis of constant-sum centipede games. These papers show that this model captures a key feature of the data, namely that as the end of the game approaches the probabilities of stopping the game increase.

Consistent with experimental research on the centipede game, research on perfect information games in general typically fails to lend support to equilibrium theories based on self-interested rational individuals with unlimited cognitive capabilities. As a result, various alternatives have been proposed. Theories of limited cognition, for instance, contend that individuals may not have unlimited computational capabilities and that they are not prone to game theoretic reasoning. Other explanations maintain that subjects may reason game theoretically, but that their preferences depend not only on their own monetary payoffs but also on those of others. In other words, these theories allow for social or payoff-interdependent preferences. For instance, Eric E. Johnson et al. (2002) investigate the extent to which limited cognition and social preferences can help explain departures from the backward induction outcome in a three-round alternating-offers bargaining game. They test for these competing explanations by conducting sessions in which players bargain with self-interested robots and by measuring patterns of information search using a computerized information display. They find that social preferences and limited cognition both play a role in detecting failures of backward induction. At the same time, they find that backward induction could be taught rapidly, although, they argue, backward induction is "simply not natural" and "presumably evolutionary adaptation did not equip people to do it." Ken Binmore et al. (2002) report experiments on one-stage and two-stage alternating-offers games, and find systematic violations of backward induction which cannot be explained by payoff-interdependent preferences. They argue that attention must turn "either to alternative formulations of preferences or to models of behavior that do not depend upon backward induction."

[6] The concept of sequential equilibrium, introduced by David M. Kreps and Robert Wilson (1982), is the main generalization of the subgame-perfect equilibrium concept to extensive games with imperfect information.

[7] For an excellent explanation of the logic behind this equilibrium see Kreps (1990, 537–43).

II. The Field Experiment

Backward induction reasoning is second nature to expert chess players. They devote a large part of their life to finding optimal strategies for innumerable chess positions using this reasoning. Further, it is common knowledge among them that they are all highly familiar with backward induction reasoning. Consequently, for two chess players playing a centipede game, it seems reasonable to think that they may *not* satisfy even the minimal departures from common knowledge of rationality that may induce rational players to depart from backward induction. Thus, Judit Polgar and Veselin Topalov, currently ranked the top female and male chess players in the world, may very well play differently from Alice and Bob in Aumann's (1992) example.[8]

In this first experiment we ask highly ranked chess players to play the one-shot version of the centipede game in an international open chess tournament. An advantage of this field setting is that it allows easy access to many highly ranked chess players. Hence, our field experiment is easier to implement than a corresponding laboratory experiment. A second advantage, at least potentially, is that a chess tournament may represent a more familiar and comfortable environment for chess players than the unfamiliar setting of a laboratory. One disadvantage, however, is the impossibility of implementing a carefully designed experiment with repetitions. This is important because without repetitions no theory can be fully rejected, because theories are predictions of steady-state behavior and not of initial responses. Nevertheless, as Costa-Gomes and Crawford (2006) emphasize, modeling initial responses more accurately promises several benefits, including obtaining insights into cognition that elucidate important aspects of strategic behavior. Thus, the results of the experiment are useful for modeling the initial responses of an interesting class of subjects, those who are likely characterized by a high degree of rationality. Further, and perhaps more importantly, initial responses that appear broadly consistent with equilibrium behavior certainly boost our confidence in such theory.

A. *Subjects*

Chess players were recruited from three international open chess tournaments in the summer of 2006 in Spain: the XXII Open International Chess Tournament of Sestao (June 17–18), the X Open International Chess Tournament of León (June 24–25), and the XXVI Open International Chess Tournament Villa de Benasque (July 6–15). In addition, we recruited subjects from the Rapid Chess Tournament of Cerler (July 10), a tournament held in conjuction with the Tournament Villa de Benasque.

Four types of players participate in a typical tournament: Grandmasters, International Masters, Federation Masters, and players with no official chess title. The title Grandmaster (henceforth GM) is awarded to world-class chess players by the World Chess Federation FIDE. It is the highest title a chess player can achieve. The title International Master (IM) ranks below the GM title, and the Federation Master (FM) is also a top title awarded by FIDE, ranking below the titles of GM and IM. In addition, all chess players are ranked according to the official Elo rating method. The difference between two players' Elo ratings is functionally related to an estimate of the probability that one of the players will beat the other should they play a chess game. The requirements for achieving a GM, IM, or FM title are somewhat complex. They involve achieving a

[8] Aumann considers a three-round (six-node) game, where the initial payoffs of \$10 and \$0.50 are multiplied by ten in each round that subjects may choose to stop. If after six rounds no player has stopped, the game ends, with both players getting zero. In his example, if there is a small ex ante probability (about 6.48×10^{-10}) that Alice *consciously and deliberately* chooses to get \$50,000 instead of \$100,000 in her last decision node, it is then rational for Bob to continue up to that point. Although this probability is very low, we would not bet on Judit Polgar making such a blunder.

prespecified Elo rating and obtaining certain outcomes in certain tournaments.[9] Typically GMs have an Elo rating above 2,500, IMs above 2,400, and FMs above 2,300. Strong club players have an Elo in the neighborhood of 1,800.

Our sample consists of 422 chess players (211 pairs): 41 GMs, 45 IMs, 29 FMs, and 307 players with no chess title. They were all recruited at the international chess tournaments at the time they were taking place. The first movers consisted of 26 GMs, 29 IMs, 15 FMs, and 141 players with no chess title. Our players with no chess title may still be considered superb chess players, as they spend several hours a week playing and studying chess, often play in regional, national, and international tournaments, and typically have a very high Elo rating. As a matter of fact, we recruited only those players with an official rating above 2,000.

For comparison purposes we also implement the same one-shot version of the centipede game with a standard pool of college student subjects in a laboratory setting. The college students were recruited from the Universidad del País Vasco in Bilbao, Spain, through campus ads and by visiting different undergraduate classes. No individual majoring in economics or mathematics was recruited. The sample consisted of 40 pairs of students. The experiments with chess players were conducted at the international open chess tournaments, while those with college students were conducted at the Universidad del País Vasco.

B. *Experimental Design*

We ran the three-round (six-node) version of the centipede game depicted in Figure 1, where the units were euros.[10] Each game involved two players who had never played the centipede game before. An experimenter read the instructions on the rules and payoffs of the game to each of the players separately, thus barring them from any opportunity to interact with each other or anyone else. Players were then placed in different rooms. Each player was informed that his opponent, who was referred to as a "player," had been read the same instructions, and that he was currently in a separate location in the same building where the experiment was taking place. Players, therefore, did not see each other and did not know each other's identity. Nonetheless, it seems reasonable to assume that in an international chess open tournament with hundreds of chess players, subjects would surmise that their opponents were also chess players. Likewise, for the students recruited through campus ads and in different undergraduate classes, it seems reasonable to assume that they believed their opponents were students.

The games were conducted through SMS messages using either a cell telephone or a BlackBerry with which the subjects entered their decisions, sent their decisions to the opponent, and received information on the decisions of the opponent. One subject was assigned the role of Player 1, and the other the role of Player 2.[11] They then participated in only one centipede game. Each player recorded his decisions and the decisions of the opponent as they occurred in a drawing of the centipede game that was similar to the figure given in the instructions. That is, players recorded the moves as they were taking place. They did not record their strategy in advance. When the game was over, each player signed his name and handed in the drawing where the joint decisions

[9] Current regulations may be found in the official FIDE Handbook (2008).

[10] At the time the experiments took place one euro was equal to 1.25 US dollars.

[11] Given the strict anonymity with which the experiment was designed, we were free to choose how to form pairs. Thus, we assigned the role of Player 1 in the Benasque tournament only to participants in that tournament, with the exception of two subjects with no chess title, and the role of Player 2 to participants in that tournament and/or in the Cerler Rapid Chess Tournament. In the Sestao and León tournaments, the role of Player 1 was assigned only to participants, again except for two subjects with no title, whereas the role of Player 2 was assigned to both participants and nonparticipants. At the aggregate level, 91.7 percent of all our sample subjects were participants in one of the four tournaments. As indicated earlier, in all cases chess players had an official Elo rating above 2,000.

Table 1—College Students: Proportion of Observations at Each Terminal Node

	N	f_1	f_2	f_3	f_4	f_5	f_6	f_7
Panel A: UPV college students								
	40	0.075	0.150	0.350	0.300	0.100	0.025	0.000
Panel B: McKelvey and Palfrey (1992) students								
Repetitions 1–5	145	0.000	0.055	0.172	0.331	0.331	0.090	0.021
Repetitions 6–10	136	0.015	0.074	0.228	0.441	0.169	0.066	0.007
Total	281	0.007	0.064	0.199	0.384	0.253	0.078	0.014

Note: The McKelvey and Palfrey students played ten repetitions of a six-node centipede game with about one-tenth lower stakes than the game played by the Universidad del País Vasco (UPV) students, who played it just once.

Table 2—Chess Players: Proportion of Observations at Each Terminal Node

Player 1	N	ELO range	f_1	f_2	f_3	f_4	f_5	f_6	f_7
Grandmasters	26	2,378–2,671	1.00	—	—	—	—	—	—
International Masters	29	2,183–2,521	0.76	0.17	0.07	—	—	—	—
Federation Masters	15	2,153–2,441	0.73	0.20	0.07	—	—	—	—
Other chess players	141	2,001–2,392	0.61	0.26	0.10	0.03	0.01	—	—
All pairs	211	2,001–2,671	0.687	0.208	0.080	0.018	0.004	—	—

had been recorded to the experimenter. Players were payed their earnings immediately after the game was played.[12]

C. *Results*

Our findings are summarized in Tables 1 and 2 and in Figures 2 and 3. They show the proportion f_i of games that ended at each of the seven possible terminal nodes $i = 1, 2, \ldots, 7$. Table 1 and Figure 2 show the results for the 40 pairs of students.

Consistent with previous experiments, we find that the large majority of players do not stop immediately. Only 3 of the 40 players who played the role of Player 1 chose to stop in the first node, while close to two-thirds of the games ended in nodes 3 and 4. For comparison, the bottom panel of the table shows the results for the college students in the McKelvey-Palfrey experiment. Although they implement the same version of the game we study, their experiment is different from ours in that they use one-tenth lower stakes and their students play ten repetitions. Nonetheless, the patterns they find are similar to ours. Even after having played several repetitions, very few students stop in the first node, and about 60 percent of their sample end in nodes 3 and 4.

Table 2 and Figure 3 show the results for each type of chess player (GM, IM, FM, and others) who take the role of Player 1. The second column reports the range of their Elo rating.

We find that the overall proportion of games that resulted in the backward induction outcome is 69 percent, almost ten times greater than the proportion of college students who made that

[12] The instructions given to the subjects can be found in a Web Appendix (http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.4.527).

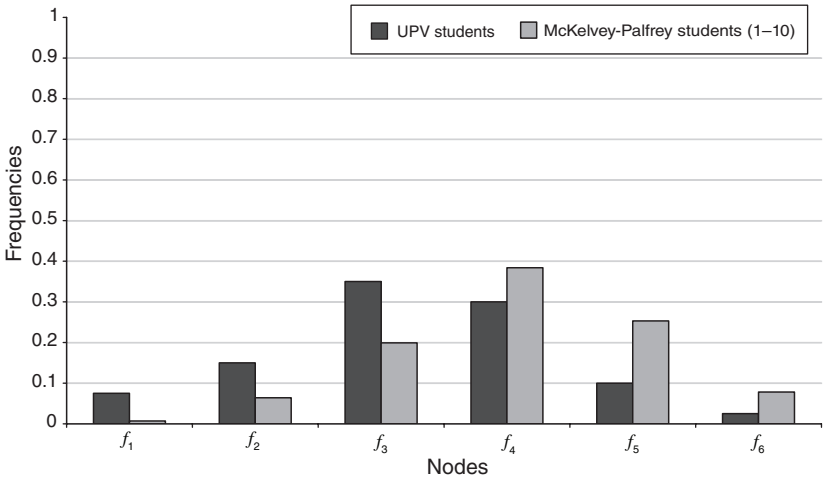

FIGURE 2. COLLEGE STUDENTS: PROPORTION OF OBSERVATIONS AT EACH TERMINAL NODE

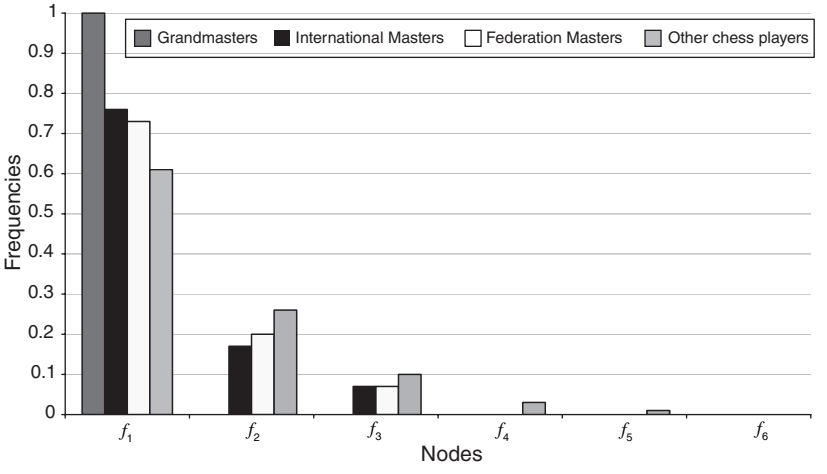

FIGURE 3. CHESS PLAYERS: PROPORTION OF OBSERVATIONS AT EACH TERMINAL NODE BY TYPE OF PLAYER 1 IN THE PAIR

choice. For the participants holding no chess titles, the proportion is 61 percent. For Federation Masters and International Masters the proportions are 73 percent and 76 percent, respectively. If we restrict our attention to Grandmasters, the proportion is a remarkable 100 percent. It is interesting to note that these proportions increase with the Elo rating of the players. A possible interpretation of this pattern is that the ideal condition of common knowledge of rationality is more closely approximated as the quality of the chess players increases.

An increase in the implied stop probabilities p_i with the rating of the players is also found for those Players 2 for whom we observe their behavior. There are 48 players with no title, 3 FMs, 10 IMs, and 5 GMs who were given the chance to take an action in node 2. Table 3 shows that the proportion that stop immediately (that is, in node 2) is 58.3 percent, 66.6 percent, 90 percent, and 100 percent, respectively.

Table 3—Chess Players: Implied Stop Probabilities at Each Terminal Node

	p_1	p_2	p_3	p_4	p_5	p_6	p_7
Grandmasters	1.00	1.00	—	—	—	—	—
	(26)	(5)					
International Masters	0.76	0.90	1.00	—	—	—	—
	(29)	(10)	(2)				
Federation Masters	0.73	0.66	1.00	—	—	—	—
	(15)	(3)	(1)				
Other chess players	0.61	0.58	0.73	0.80	1.00	—	—
	(141)	(48)	(19)	(5)	(1)		

Note: The number of players observed making a decision (stop or continue) at each node is in parentheses.

The main conclusions that we can draw from our field experiments are that (i) chess players tend to play very differently from college students, and that (ii) a significant majority of chess players chose the only action that is consistent with equilibrium.[13] These results are consistent with the idea that chess players represent a unique subject pool with many levels of mutual knowledge of rationality. Further, the fact that their initial responses are so close to equilibrium certainly boosts our confidence in a theory that gives a central role to the principles of self-interested rational economic agents and to their assessments of the rationality of their opponents. Motivated by these findings, we turn next to our main experiment.

III. The Laboratory Experiment

The objective of this experiment is to study whether players' assessments of their opponents' rationality is a key determinant of whether the subgame-perfect equilibrium is a good predictor of behavior. The experiment takes place in a laboratory setting where we match both chess players and students with either chess players or students in four different treatments, depending on the order of play. The treatment where we match students with students is useful simply to replicate the main results obtained in previous experiments. The treatment where we have chess players facing chess players is a complement of the initial field experiment studied earlier since, by allowing learning and experimentation, one can observe whether chess players converge to the equilibrium outcome. The two treatments where we have students face chess players are the most important ones. The fact that most people should not be surprised that chess players are good at backward induction and that, indeed, as evidenced by the previous section, they tend to play according to it, is what renders the matching between students and chess players a powerful tool. If knowledge of an opponent's rationality is an important determinant of one's behavior, then students should alter their behavior compared to the situation where they face another student. Likewise, to the extent that chess players may be less confident on the rationality of students than on the one displayed by other chess players, they should also alter their behavior relative to the situation where they face another chess player.[14]

[13] Equilibrium predictions are about stationary situations, and not about initial responses. Thus, not surprisingly, the equilibrium strategies are not best responses to the observed behavior. Player 1's best response to the population frequencies is to continue in the first two nodes, and Player 2's best response is to continue in his first node and to stop in his second node.

[14] Although chess players conform rather closely to the equilibrium predictions in the field experiment, it is certainly possible that they were playing a different game than the one the experimenter has created. Perhaps they do not intend so much to maximize their monetary reward as to "beat" their opponent. That is, chess players may like to win, and

A. *Subjects*

College students were recruited from the Universidad del País Vasco in Bilbao, Spain. None of the participating students was majoring in mathematics or economics. Chess players were recruited from a number of chess clubs from the Bilbao area affiliated with the Spanish Chess Federation. None of the players had an official chess title, and their average Elo rating is 2,007, ranging from 1,817 to 2,205. That is, their ratings are in the range of the lowest ranked chess players who participated in the field experiment studied in the previous section.

B. *Experimental Design*

The experimental design is very similar to that in McKelvey and Palfrey (1992). Each experiment consisted of two sessions of ten repetitions of the centipede game depicted in Figure 1. In each session, 20 subjects, none of whom had previously played a centipede game, were divided into two equally sized groups which we called the white group and the black group. White players played the role of Player 1, and black players played the role of Player 2. Each white (black) player played one instance of the centipede game with each one of the black (white) players, without knowing his identity. No subject participated in more than one session. We followed McKelvey and Palfrey's design as much as possible, including their matching algorithm which is meant to prevent supergame or cooperative behavior. We deviate from their design in that we used the same payoffs we used in the field, which, after adjusting for inflation, are about ten times larger than theirs, and in that after the instructions were read, players were located in individual rooms with no visual contact with each other. As in the field experiment, players sent their move choices through SMS messages rather than through computer terminals.[15]

The only feature that differentiates the four experiments is the nature of the pool of subjects in each of the groups. In Treatment I, both groups consisted of college students. In Treatment II, the white group consisted of college students and the black group of chess players. In Treatment III, the white group was composed of chess players and the black group of college students. Finally, in Treatment IV, chess players faced chess players. Most importantly, in all the treatments the composition of the two groups (though not the identity of its members) was common knowledge among the players.[16] The sessions were conducted at the Universidad del País Vasco in February 2007. Table 4 summarizes the experimental design.

C. *Results*

Table 5, panel A, shows the proportion of games in each session that ended up in each of the seven possible terminal nodes, and panel B reports the implied probabilities of stopping, conditional on having reached a given node.

As can be observed, when students play against other students, the distribution of observations resembles that in previous experiments of similar six-node exponential centipede games.

one way to win is to obtain a higher payoff than the opponent. Another alternative could be that chess players cannot allow themselves to give an "incorrect" answer to a (chess) puzzle, no matter how much money they lose by doing so. These alternatives suggest that chess players should *not* alter their behavior much when facing a student relative to the situation when they play another chess player.

[15] The instructions given to the players can be found in the Web Appendix.

[16] To further preclude the possibility of cooperation, we made certain that no students of the same entering class and major played in different groups in treatment I, that no chess players belonging to the same chess club or that had participated in the same tournament in the last four years played in different groups in treatment IV, and that chess players participating in treatments II and III were not college students.

Table 4—Experimental Design for Laboratory Experiment

Treatment	Subject pool Player 1 (white)	Subject pool Player 2 (black)	Session	Subjects	Games per subject	Total games
I	Students	Students	1	20	10	100
			2	20	10	100
II	Students	Chess players	3	20	10	100
			4	20	10	100
III	Chess players	Students	5	20	10	100
			6	20	10	100
IV	Chess players	Chess players	7	20	10	100
			8	20	10	100

Very few subjects (3 percent) stop immediately, and over 60 percent stop at nodes 3 or 4.[17] The way that students play, however, drastically changes when they are informed that they are playing against chess players. When they take up the role of Player 1 (Treatment II), the proportion of observations ending in terminal node 1 (30 percent) is ten times greater than when they play against a student, and even after two moves the implied stop probability, 0.61, is 50 percent greater than when they play against students, 0.42. Likewise, when they take up the role of Player 2 (Treatment III) the distribution of games across the resulting terminal nodes is stochastically dominated by the distribution corresponding to the first treatment.

The main observation one can infer from these results is that college students' behavior depends on whether they face a highly rational opponent or a fellow student. This dependence raises the question of whether students are unaware of backward induction reasoning. It seems that they may or may not subscribe to such reasoning depending on their beliefs about the assessed sophistication and experience of their opponent.

We now turn our attention to the chess players. First, we find that when they play against other chess players the aggregate distribution of observations is not much different from what we found in the field: about 70 percent of the games end immediately. Yet, chess players, like the students, play drastically differently when told that they are playing against a student. The proportion of observations ending in the first node in Treatment IV is almost twice that observed in Treatment III, and the implied stop probabilities are greater in every node in Treatment IV relative to the case when they play against a student (nodes 1 and 3 in Treatment III, and nodes 2 and 4 in Treatment II).

The differences in stop probabilities are such that the distributions of the proportion of observations in both Treatments II and III are stochastically dominated by that in Treatment I, while the distribution in Treatment IV is dominated by those in Treatments II and III. Comparing the latter two treatments, chess players have a greater implied stop probability than students in three of the first four nodes, and the implied stop probabilities tend to increase monotonically with the stage of the game in every treatment and, for Treatments II and III, also for a given type of player.[18]

Table 6 disaggregates the data into "early" plays (games 1–5) and "late" plays (games 6–10).

Consistent with past experiments, we find that for each treatment the distribution of observations in the early plays stochastically dominates that in the late plays. As in the aggregate data, implied stop probabilities tend to increase as we get closer to the last move in each of the

[17] Perhaps not surprisingly, as we use much greater payoffs than in past experiments, the distribution is slightly to the left of the corresponding McKelvey-Palfrey (1992) distribution.

[18] The one possible exception to this pattern is the second node in the treatment IV, although as it will be noted later this is actually the result of aggregating across rounds with very different stop probabilities.

TABLE 5—PROPORTION OF OBSERVATIONS AND IMPLIED STOP PROBABILITIES AT EACH TERMINAL NODE

	Session	N	f_1	f_2	f_3	f_4	f_5	f_6	f_7
Panel A: Proportion of observations f_i									
I. Students versus students	1	100	0.04	0.15	0.40	0.27	0.13	0.01	0
	2	100	0.02	0.18	0.28	0.33	0.14	0.04	0.01
	Total 1–2	200	0.030	0.165	0.340	0.300	0.135	0.025	0.005
II. Students versus chess players	3	100	0.28	0.36	0.19	0.11	0.06	0	0
	4	100	0.32	0.37	0.22	0.07	0.02	0	0
	Total 3–4	200	0.300	0.365	0.205	0.090	0.040	0	0
III. Chess players versus students	5	100	0.37	0.26	0.22	0.09	0.06	0	0
	6	100	0.38	0.29	0.17	0.10	0.06	0	0
	Total 5–6	200	0.375	0.275	0.195	0.095	0.060	0	0
IV. Chess players versus chess players	7	100	0.69	0.19	0.11	0.01	0	0	0
	8	100	0.76	0.16	0.07	0.01	0	0	0
	Total 7–8	200	0.725	0.175	0.090	0.010	0	0	0

	Session	p_1	p_2	p_3	p_4	p_5	p_6
Panel B: Implied stop probability p_i							
I. Students versus students	1	0.04	0.16	0.49	0.66	0.93	1.00
		100	96	81	41	14	1
	2	0.02	0.18	0.35	0.63	0.74	0.80
		100	98	80	52	19	5
	Total 1–2	0.03	0.17	0.42	0.65	0.82	0.83
		200	194	161	93	33	6
II. Students versus chess players	3	0.28	0.50	0.53	0.65	1.00	—
		100	72	36	17	6	0
	4	0.32	0.54	0.71	0.78	1.00	—
		100	68	31	9	2	0
	Total 3–4	0.30	0.52	0.61	0.69	1.00	—
		200	140	67	26	8	0
III. Chess players versus students	5	0.37	0.41	0.59	0.60	1.00	—
		100	63	37	15	6	0
	6	0.38	0.47	0.52	0.63	1.00	—
		100	62	33	16	6	0
	Total 5–6	0.375	0.44	0.56	0.61	1.00	—
		200	125	70	31	12	0
IV. Chess players versus chess players	7	0.69	0.61	0.92	1.00	—	—
		100	31	12	1	0	0
	8	0.76	0.67	0.88	1.00	—	—
		100	24	8	1	0	0
	Total 7–8	0.725	0.64	0.90	1.00	—	—
		200	55	20	2	0	0

Note: In panel B, the number of players making a decision at each node is indicated below the implied stop probabilities.

treatments, for both early and late plays. The only exception to this pattern occurs in the last node of the late plays in the first treatment, where the probability drops from 0.76 to 0.66. But these probabilities are based on only three subjects, one of whom decided to continue rather than stop in the sixth node.

TABLE 6—PROPORTION OF OBSERVATIONS AND IMPLIED STOP PROBABILITIES IN EARLY ("1–5") AND LATE ("6–10") GAMES AT EACH TERMINAL NODE

Treatment	Games	N	f_1	f_2	f_3	f_4	f_5	f_6	f_7
Panel A: Proportion of observations f_i									
I. Students versus students	"1–5"	100	0.01	0.06	0.37	0.36	0.17	0.03	0
	"6–10"	100	0.05	0.27	0.31	0.24	0.10	0.02	0.01
II. Students versus chess players	"1–5"	100	0.13	0.41	0.21	0.17	0.08	0	0
	"6–10"	100	0.47	0.32	0.20	0.01	0	0	0
III. Chess players versus students	"1–5"	100	0.15	0.32	0.24	0.17	0.12	0	0
	"6–10"	100	0.60	0.23	0.15	0.02	0	0	0
IV. Chess players versus chess players	"1–5"	100	0.45	0.35	0.18	0.02	0	0	0
	"6–10"	100	1.00	0	0	0	0	0	0

	Games	p_1	p_2	p_3	p_4	p_5	p_6	p_7
Panel B: Implied stop probability p_i								
I. Students versus students	"1–5"	0.01	0.06	0.40	0.64	0.85	1.00	—
		100	99	93	56	20	3	0
	"6–10"	0.05	0.28	0.46	0.64	0.76	0.66	1.00
		100	95	68	37	13	3	1
II. Students versus chess players	"1–5"	0.13	0.47	0.46	0.68	1.00	—	—
		100	87	46	25	8	0	0
	"6–10"	0.47	0.60	0.95	1.00	—	—	—
		100	53	21	1	0	0	0
III. Chess players versus students	"1–5"	0.15	0.38	0.45	0.58	1	—	—
		100	85	53	29	12	0	0
	"6–10"	0.60	0.58	0.88	1.00	—	—	—
		100	40	17	2	0	0	0
IV. Chess players versus chess players	"1–5"	0.45	0.64	0.90	1.00	—	—	—
		100	55	20	2	0	0	0
	"6–10"	1.00	—	—	—	—	—	—
		100	0	0	0	0	0	0

Note: In panel B, the number of players making a decision at each node is indicated below the implied stop probabilities.

Treatments II and III show that when students and chess players play against each other, they do not behave very differently from each other. Although chess players tend to have a greater implied stop probability at every node, the magnitude of the differences is not large, and in both treatments the probability of stopping reaches one in node 5 in the early plays and in node 4 for the late plays.

As in the aggregate data, both for early and late plays, the distributions of observations in treatments II and III are stochastically dominated by that in Treatment I, while they dominate the distribution corresponding to Treatment IV. More important, all the late games of Treatment IV ended at the first terminal node.[19] This result indicates that chess players need just a small

[19] The fact that in the late games all subjects stop in the first node explains why in Table 5, panel B, the implied stop probabilities, where the data are aggregated over all games, decrease from node 1 to node 2.

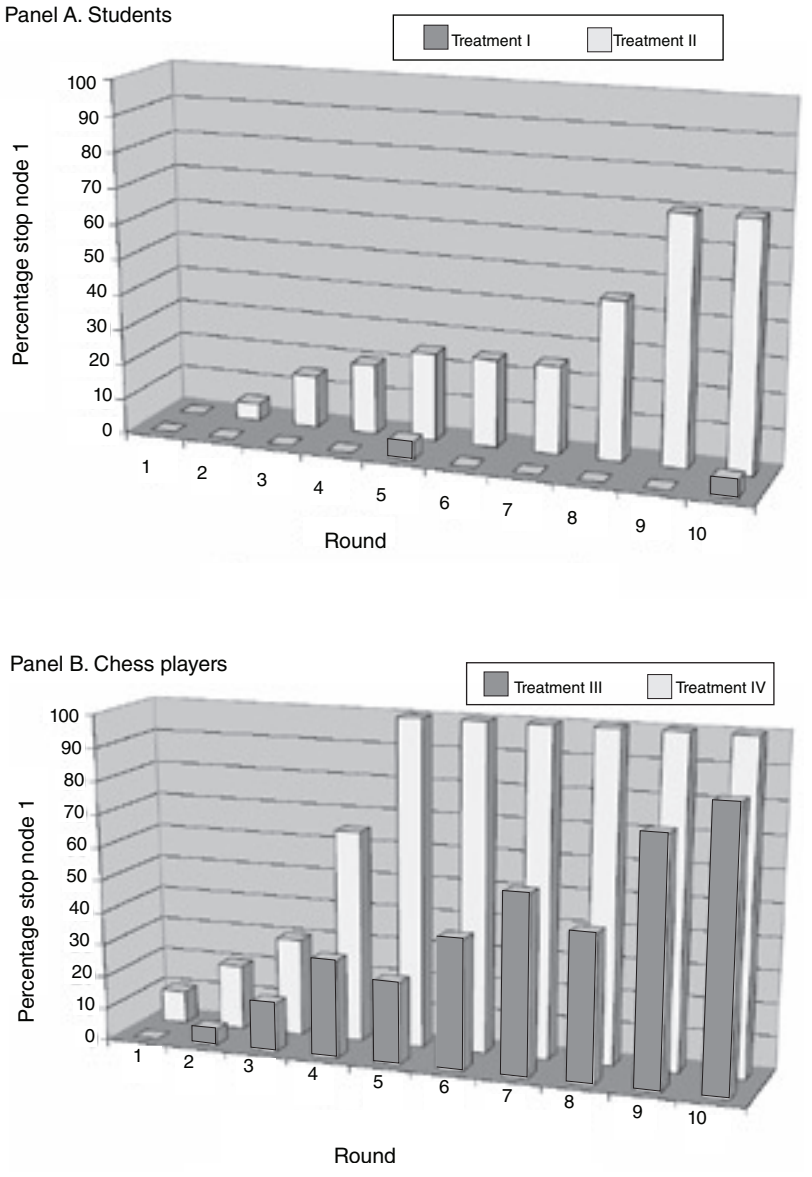

FIGURE 4. PERCENTAGE OF "STOP" IN NODE 1 PER ROUND

number of repetitions to learn to predict other chess players' behavior correctly and to converge to equilibrium. Their behavior, therefore, is not inconsistent with the hypothesis that they satisfy the condition of common knowledge of rationality.

Finally, Figure 4 reports the proportion of games that ended in the first node at each round and for each treatment. Panel A represents the behavior of students and panel B the behavior of chess players.

These round-by-round data show in more detail the reactions of college students and chess players to the different types of opponents they face.[20] The evidence from Treatments II and III reveals that they are not very different from each other. More important, in rounds 9 and 10 of Treatment II, 70 percent of students stop immediately, whereas in the same rounds of Treatment III, 75 and 85 percent of chess players stop immediately. Hence, these mixed treatments show a substantial degree of convergence toward equilibrium.

It is interesting that chess players playing against chess players seem to "experiment" during the first few repetitions by choosing to "continue" much more frequently than when playing a one-shot game in the field. In panel B, the proportion of Treatment IV games that ended at the first node steadily increases from 10 percent in the first games to 100 percent in the fifth repetition. Hence, although the aggregate distribution reported in panel A of Table 5 is similar to the distribution obtained in the field, chess players drastically alter their behavior in the laboratory initially when they know they will play ten repetitions of the same game, and then they all quickly converge to equilibrium.

IV. Conclusions

Aumann (1998) showed that if the backward induction outcome is not played at some state of the world, then at that state there must be a node in the path of play at which the player whose turn it is to move deliberately chooses an action that he *knows* yields him a lower payoff than the one he would get by choosing an alternative action. Specifically, at that state there is a node that is reached along the path of play at which a player chooses to continue, even though he knows at the time of his choice that stopping is more profitable. Although this irrational behavior is by no means impossible among humans, our working hypothesis is that it is less likely to occur among chess players, who are familiar with backward induction reasoning. Further, their familiarity with this form of reasoning is common knowledge among many, if not all, humans.

In this paper, we have used chess players in two experiments. Our first experiment takes place in a field setting where we elicit only their "initial responses," that is where we study their strategic thinking having suppressed learning and repeated game effects. We find that even at the level of initial responses, chess players' behavior is remarkably close to equilibrium. Our laboratory experiment then lends conclusive support to the equilibrium hypothesis by further showing that chess players, when allowed minimal opportunities to experiment and learn, converge very rapidly to equilibrium behavior. These results suggest that the "ideal" condition of common knowledge of rationality seems to be approached closely when chess players play the centipede game.

Our main findings concern the standard pool of subjects in the laboratory experiments. In games that involved one college student facing one chess player, the backward induction outcome occurred more than ten times more often than in games involving college students only, and already by the tenth repetition college students approached it quite closely. We view these findings as being highly consistent with the predictions of the theoretical literature. It is the rationality of a subject and his assessment of the opponent's rationality, rather than altruism or other forms of social preferences, that seem to be key to predicting the outcome of perfect information games. Thus, in the context of the extensive recent discussion in the literature about nonstandard assumptions on players' preferences as a realistic and necessary modeling device, this paper suggests that such assumptions might be neither. With respect to future research, our findings

[20] The fact that chess players play very differently when matched with other chess players than when matched with students also means that they are not simply trying to beat their opponent by obtaining a higher payoff or that they do not require themselves to give "the correct answer" to what they could perceive as a chess puzzle.

can also be interpreted as representing a sensible shift away from limited cognition and learning backward induction toward deciding when to apply equilibrium theory.

REFERENCES

Asheim, Geir B., and Martin Dufwenberg. 2003. "Deductive Reasoning in Extensive Games." *Economic Journal*, 113(487): 305–25.

Aumann, Robert J. 1992. "Irrationality in Game Theory." In *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*, ed. Partha Dasgupta, Douglas Gale, Oliver Hart, and Eric Maskin, 214–27. Cambridge, MA: MIT Press.

Aumann, Robert J. 1995. "Backward Induction and Common Knowledge of Rationality." *Games and Economic Behavior*, 8(1): 6–19.

Aumann, Robert J. 1998. "On the Centipede Game." *Games and Economic Behavior*, 23(1): 97–105.

Ben-Porath, Elchanan. 1997. "Rationality, Nash Equilibrium and Backwards Induction in Perfect-Information Games." *Review of Economic Studies*, 64(1): 23–46.

Binmore, Ken, John McCarthy, Giovanni Ponti, and Larry Samuelson. 2002. "A Backward Induction Experiment." *Journal of Economic Theory*, 104(1): 48–88.

Bornstein, Gary, Tamar Kugler, and Anthony Ziegelmeyer. 2004. "Individual and Group Decisions in the Centipede Game: Are "Groups" More Rational Players?" *Journal of Experimental Social Psychology,* 40(5): 599–605.

Costa-Gomes, Miguel A., and Vincent P. Crawford. 2006. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study." *American Economic Review*, 96(5): 1737–68.

Fey, Mark, Richard D. McKelvey, and Thomas R. Palfrey. 1996. "An Experimental Study of Constant-Sum Centipede Games." *International Journal of Game Theory*, 25(3): 269–87.

FIDE Handbook. 2008. http://www.fide.com/info/handbook (accessed August 27, 2008).

Johnson, Eric J., Colin Camerer, Sankar Sen, and Talia Rymon. 2002. "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining." *Journal of Economic Theory*, 104(1): 16–47.

Kreps, David M. 1990. *A Course in Microeconomic Theory.* Princeton: Princeton University Press.

Kreps, David M., and Robert Wilson. 1982. "Sequential Equilibria." *Econometrica*, 50(4): 863–94.

McKelvey, Richard D., and Thomas R. Palfrey. 1992. "An Experimental Study of the Centipede Game." *Econometrica*, 60(4): 803–36.

McKelvey, Richard D., and Thomas R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior,* 10(1): 6–38.

McKelvey, Richard D., and Thomas R. Palfrey. 1998. "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics*, 1(1): 9–41.

Nagel, Rosemarie, and Fang Fang Tang. 1988. "Experimental Results on the Centipede Game in Normal Form: An Investigation on Learning." *Journal of Mathematical Psychology,* 42(2): 256–84.

Rapoport, Amnon, William E. Stein, James E. Parco, and Thomas E. Nicholas. 2003. "Equilibrium Play and Adaptive Learning in a Three-Person Centipede Game." *Games and Economic Behavior*, 43(2): 239–65.

Reny, Philip J. 1992. "Rationality in Extensive-Form Games." *Journal of Economic Perspectives*, 6(4): 103–18.

Reny, Philip J. 1993. "Common Belief and the Theory of Games with Perfect Information." *Journal of Economic Theory*, 59(2): 257–74.

Rosenthal, Robert W. 1981. "Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox." *Journal of Economic Theory*, 25(1): 92–100.

Zauner, Klaus G. 1999. "A Payoff Uncertainty Explanation of Results in Experimental Centipede Games." *Games and Economic Behavior*, 26(1): 157–85.