Field Centipedes*

Ignacio Palacios-Huerta† Oscar Volij‡

November 2006

Abstract

We conduct a field experiment in which highly-ranked chess players play the centipede game in a natural setting. In the experiment two players alternately are faced with the decision of either taking an exponentially growing pile of money and ending the game, or letting the other player make the decision. The player who decides to stop the game takes the larger portion of the pile, and the other player gets the remaining amount. All standard equilibrium concepts dictate that the player who decides first must stop the game immediately. There is vast experimental evidence, however, that this rarely occurs. Contrary to this evidence our results show that 69% of chess players stop the game immediately. When we restrict attention to chess Grandmasters this percentage escalates to 100%.

Keywords: Backward Induction, Centipede Game, Field Experiments, Rationality.

*We are grateful to the organizers of the XXII Open International Chess Tournament of Sestao, the X Open International Chess Tournament of Leon, and the XXVI Open International Chess Tournament Villa de Benasque for access to the players that participated in this study. Comments and suggestions by Jose Apesteguía, Pedro Dal Bó, Leontxo García and Yona Rubinstein are appreciated. Financial support from the Salomon Foundation and the Spanish Ministerio de Ciencia y Tecnología (grant BEC2003-08182) is gratefully acknowledged.

†Brown University. Email: ipalacios@brown.edu.

‡Iowa State University and Ben-Gurion University. Email: volij@iastate.edu.

1 Introduction

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.

One instance of the centipede game is as follows. A pile of $4 and a pile of $1 are lying on a table. Player I has two options, either to "stop" or to "continue." If he stops, the game ends and he gets $4 while Player II gets the remaining dollar. If he continues, the two piles are doubled, to $8 and $2, and Player II 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 I, or to let the piles double again and let Player I decide. The game continues for at most six periods. If by then neither of the players have stopped, Player I gets $256 and Player II 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 I will stop at the first opportunity, getting just $4. Despite this unambiguous prediction, however, game theorists often "wonder if it really reflects the way in which *anyone* would play such a game" (McKelvey and Palfrey, 1992, italics added).

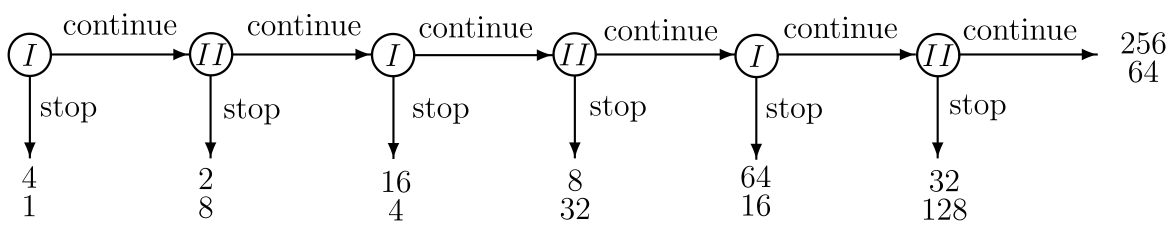

Figure 1: A centipede game.

The game theoretic prescription for this kind of sequential games goes so much against people's intuition that it induced 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] 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."

The apparent conflict between the theoretical prediction and intuitively reasonable behavior 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 periods. Reny (1992) also eloquently demonstrates how violating backward induction may be perfectly rational. Ben-Porath (1997) shows that several rounds of "continuation" are consistent with *common certainty* of rationality. Therefore, rationality alone does not imply the pessimistic and pretty unprofitable behavior prescribed by the backward inductive outcome.

It actually turns out that it is not rationality, or 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.[2] However, he also concedes that common knowledge of rationality "is an ideal condition that is rarely met in practice" (p.18), and that absent this condition, 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.

Uneasiness with the backward induction outcome arose long before the first experimental study of the centipede game took place. 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% of the games played in McKelvey-Palfrey's centipede game

[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] 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.

experiment resulted in the backward induction outcome, even after subjects played several repetitions of the game, and these findings have been confirmed by other studies.[3]

There were some later attempts to experimentally test the backward induction prediction in centipede-like games. Since the pie to be divided between the players in these games grows as play advances to later nodes, one could explain the tendency not to exit at early nodes by means of a small measure of altruism. 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.[4] Indeed, when they ran two kinds of constant-sum centipede games, one with ten nodes and a second one with six nodes, the proportion of games that resulted in the backward induction outcome was 45% and 59% respectively. Although this is a dramatic increase in the performance of the theoretical prediction, Fey, McKelvey and Palfrey (1996) still concluded that backward induction is inadequate for explaining players' behavior.

Another attempt at achieving the backward induction outcome was recently implemented by 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. Here again, the backward inductive outcome was observed to be played more often that in McKelvey-Palfrey's experiment, 46% of the trials, but nonetheless was not enough to support the theoretical predictions.

The line of research just described shows that it is possible to find versions of

[3]For instance, Nagel and Tang (1998) implement an experiment on the centipede game played in reduced normal form. Even after subjects repeat the game one hundred times against randomly selected opponents, fewer than 1% of the games end in the backward induction outcome. Bornstein et al. (2004) find in their sample that even if individuals play in groups no games end in any of the first two nodes.

[4]By moderate altruism we mean other-regarding preferences that still consider a dollar to oneself as preferred to the same dollar to the other.

the centipede game (constant-sum rather than exponential) and variations of how the game is implemented (three rather than two-person with higher payoffs and more repetitions) where the backward induction prediction performs better than in McKelvey and Palfrey (1992), although still far from perfectly. In this paper we do not consider these variations. Rather, we intentionally choose a version of the centipede game where the typical pool of experimental subjects performs at its worst, namely, the exponential game. We also consider arguably low payoffs. Our first departure from the typical experiment is that we allow no repetitions. Our subjects play the game only once, thus eliminating the learning opportunities which past experiments have shown to facilitate the backward induction outcome. Our second departure is more drastic. Motivated by the recent literature on field experiments, we do not employ the standard pool of experimental subjects (college students) in a laboratory setting.[5] Instead, we ask highly-ranked chess players to play the centipede game in a setting that is familiar to them: an international open chess tournament. Although these players are not familiar with the centipede game, backward induction reasoning is second nature to them. Indeed they devoted a large part of their life finding optimal strategies to innumerable chess positions using this reasoning. More importantly, one can safely say that it is common knowledge among chess players that they are all highly familiar with backward induction reasoning. Consequently, for two chess players playing a centipede game, it 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, Judith Polgar and Veselin Topalov, currently ranked the top female and male players in the world, may very well play differently from Alice and Bob in Aumann's (1992) example.[6]

Besides implementing the centipede game for chess players, we also implement it for a standard pool of college student subjects in a laboratory setting.

Our findings show that, consistent with previous results in the experimental literature, only a small percentage of students (7.5%) play the backward induction

[5]The literature on field experiments is vast and growing rapidly. See Harrison and List (2004) for a comprehensive survey.

[6]Aumann considers a three round (six period) game, where the initial payoffs $10 and $0.50 are multiplied by 10 each period that subjects may choose to stop. If after six rounds no player has stopped, the game ends, with both players getting 0. 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 Judith Polgar making such a blunder.

outcome. Chess players, however, behave drastically different: 69% of them stop the game immediately. When we restrict attention to chess Grandmasters, this percentage escalates to 100%.

2 Experimental Procedures

2.1 Subjects

Chess players were recruited from the participants in three international open chess tournaments that have taken place 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 Spain. 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 masters by the World Chess Federation FIDE. It is the highest title a chess player can attain. 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 ELO rating method. The difference between two players' ELO ratings is functionally related to an estimate of the probability that one of the players beat the other if they played a chess game. The requirements for achieving a GM, IM or FM title are somewhat complex. They involve achieving a pre specified ELO rating and obtaining certain outcomes in certain tournaments.[7] Typically GMs have an ELO rating above 2500, IMs above 2400 and FMs above 2300.

Our sample consists of 211 pairs of chess players. 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, in many dimensions, still be considered superb chess players, as they spend several hours per week playing and studying chess, often play in regional, national and international tournaments, and they typically have a very high ELO rating. Indeed, in our sample they all have a rating above 2,000.

In addition to the chess players, we also recruited forty pairs of college students

[7] Current regulations may be found in the official FIDE Handbook.

from the Universidad del País Vasco in Bilbao (Spain) through campus ads and by visiting different undergraduate classes. No individual majoring in economics and mathematics was recruited.

The experiments with chess players were conducted at international open chess tournaments, while the experiments with college students were conducted at the Universidad del Pais Vasco.

2.2 Experimental Design

We ran the three round (six-move) version of the centipede game described in Section 1, where the units are euros.[8] Each game involves 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, 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 had been read the same instructions, and that he was currently in a separate location. Players, therefore, did not see each other and did not know each other's identity. Still, in an international chess open tournament with hundreds of chess players, it seems reasonable to assume that chess players surmised 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 which the subjects used to enter their decisions, send their decisions to the opponent, and receive information on the decisions of the opponent. One subject was assigned the role of Player I, and the other the role of Player II. They then participated in only one centipede game. Each player recorded his decisions and the decisions of the opponent in a drawing of the centipede game that was similar to the figure given in the instructions. When the game was over each player signed his name and handed in the drawing where the joint decisions had been recorded to the experimenter. Players were payed their earnings immediately after the game was played. The actual instructions given to the subjects are found at http://volij.co.il/publications/papers/Centipede_Instructions_Sestao.pdf.

[8]At the time the experiments took place 1 euro = 1.25 US dollars.

3 Results

Our findings are summarized in Tables 1 and 2, and Figures 2 and 3. They show the proportion f_n of games that ended at each of the seven possible terminal nodes $n = 1, 2, \ldots, 7$.

First, Table 1 and Figure 2 show the results for the 40 pairs of students.

[Table 1 here]

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 I chose to stop in the first node, while two thirds of the games end in nodes f_3 and f_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 that we study, their experiment is different from ours in that they use one tenth lower stakes and their students play ten repetitions. The patterns they find, however, are similar to ours. Even after having played several repetitions, very few students stop in the first node and about two thirds of their sample end in nodes f_3 and f_4.

[Figure 2 here]

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

[Table 2 here]

We find that the overall proportion of games that resulted in the backward induction outcome is 69%, almost ten times greater than the proportion of college students that make that choice. For the participants holding no chess titles the proportion is 61%. For Federation Masters and International Masters the proportions are 73% and 76%, respectively. If we restrict our attention to Grandmasters, the proportion is a remarkable 100%.

[Figure 3 here]

4 Conclusions

Aumann (1998) showed that if the backward induction outcome is not played, then there must be a node in the path of play at which the player whose turn 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, 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. Our experimental results are consistent with this hypothesis. Even in the six-node, exponential version, the backward induction outcome occurred almost ten times more often in the games played by chess players than in the ones played by college students. The "ideal" condition of common knowledge of rationality seems to be approached closely when chess players play the centipede game, especially when we restrict attention to chess Grandmasters.

REFERENCES

Aumann, R. J. (1992), "Irrationality in Game Theory," in P. Dasgupta et al (Ed.), *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*, pp. 214-27. MIT Press.

Aumann, R. J. (1995), "Backward Induction and Common Knowledge of Rationality," *Games and Economic Behavior* 8, 6-19.

Aumann, R. J. (1998), "On the Centipede Game," *Games and Economic Behavior* 23, 97-105.

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

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

Fey, M., R. D. McKelvey, and T. R. Palfrey (1996), "An Experimental Study of Constant-Sum Centipede Games," *International Journal of Game Theory* 25, 269-87.

FIDE Handbook, http://www.fide.com/official/handbook.asp.

Harrison, G. W. and J. A. List (2004), "Field Experiments," *Journal of Economic Literature* 42(4), 1009-1055.

McKelvey, R. D. and T. R. Palfrey (1992), "An Experimental Study of the Centipede Game," *Econometrica* 60, 803-836.

Nagel, R., and F. F. Tang (1988), "Experimental Results on the Centipede Game in Normal Form: An Investigation on Learning," *Journal of Mathematical Psychology* 42, 356-384.

Rapoport, A., W. E. Steinb, J. E. Parcoc, and T. E. Nicholas (2003), "Equilibrium Play and Adaptive Learning in a Three-Person Centipede Game," *Games and Economic Behavior* 43(2), 23965.

Reny, P. J. (1992), "Rationality in Extensive Form Games," *Journal of Economic Perspectives* 6(4), 103-118.

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

Rosenthal, R. W. (1981), “Games of Perfect Information, Predatory Pricing and Chain Store Paradox,” *Journal of Economic Theory* 25(1), 92-100.

TABLE 1 – COLLEGE STUDENTS

PROPORTIONS OF OBSERVATIONS AT EACH TERMINAL NODE f_i

	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

TABLE 2 – CHESS PLAYERS

PROPORTIONS OF OBSERVATIONS AT EACH TERMINAL NODE f_i

	N	*ELO range*	f_1	f_2	f_3	f_4	f_5	f_6	f_7
Grandmasters	26	2378-2671	1.00	0.00	0.00	0.00	0.00	0.00	0.00
International Masters	29	2183-2521	0.76	0.17	0.07	0.00	0.00	0.00	0.00
Federation Masters	15	2153-2441	0.73	0.20	0.07	0.00	0.00	0.00	0.00
Other chess players	141	2001-2392	0.61	0.26	0.10	0.03	0.01	0.00	0.00

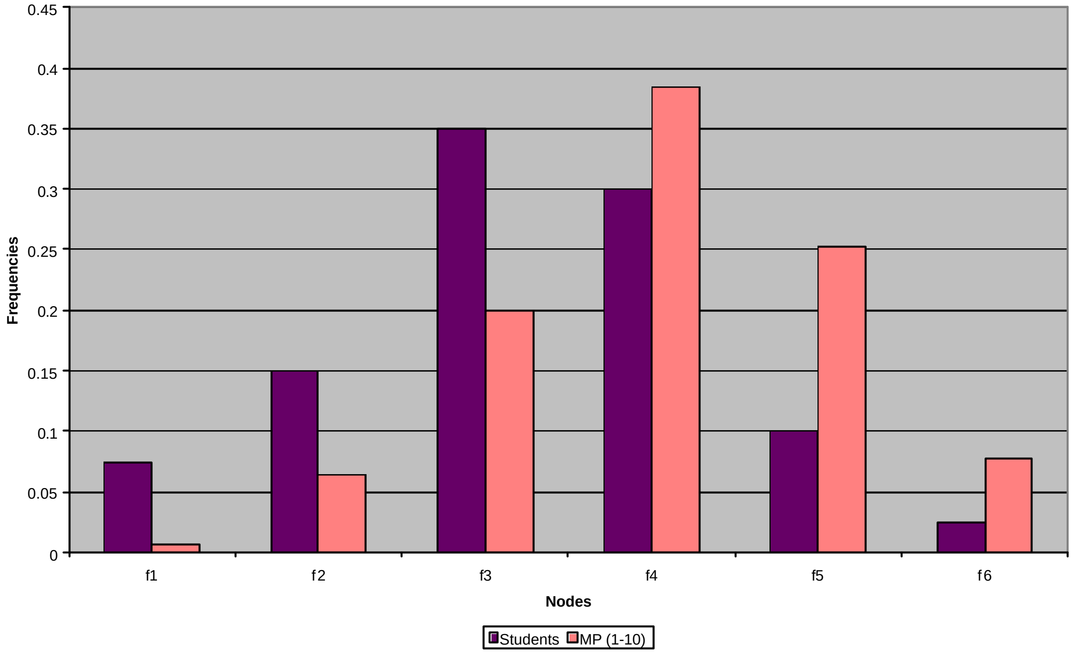

Figure 2: College Students
Relative frequencies of stops at each terminal node

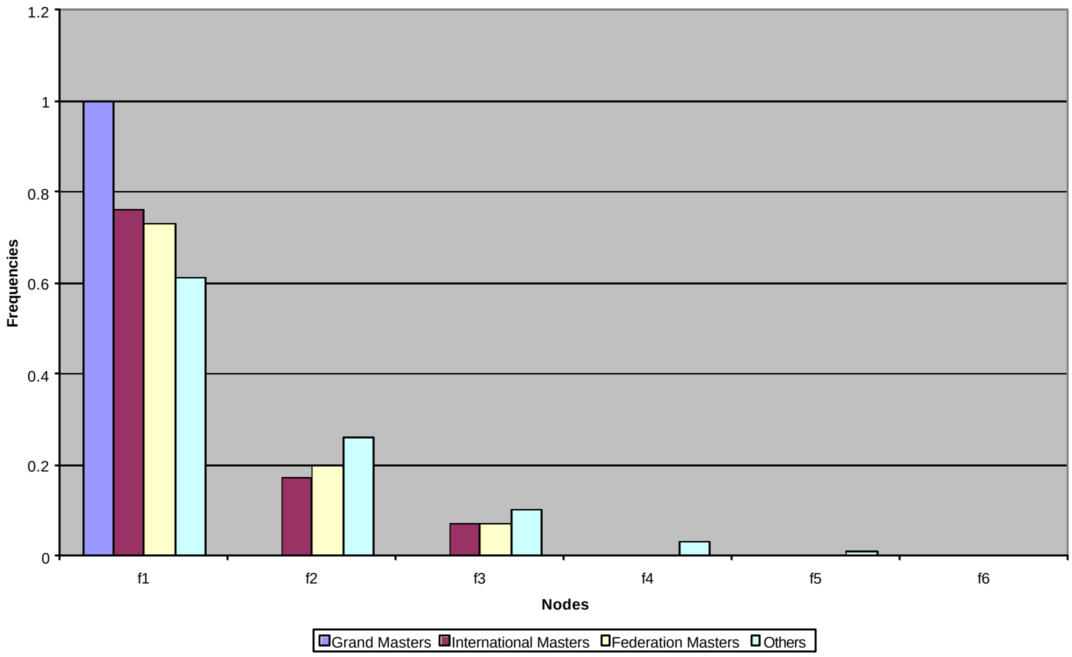

Figure 3: Chess Players
Relative frequencies of stops at each terminal node