What Happens in the Field Stays in the Field:
Professionals Do Not Play Minimax in Laboratory Experiments

Steven D. Levitt, John A. List, and David H. Reiley*

August 2007

Abstract

　　The minimax argument represents game theory in its most elegant form: simple but with stark predictions. Although some of these predictions have been confirmed in field data, tests of minimax have generally failed in laboratory experiments that make use of student subjects. In a seminal study, Palacios-Huerta and Volij (2007) present evidence that potentially resolves this puzzle: soccer players (both amateur and professional) play minimax strategies in laboratory experiments. They attribute this result to an ability on the part of the soccer players to transfer expertise they have learned shooting penalty kicks to card games played in the lab. These results induced us to carry out similar experiments using subject pools with unrivaled experience applying analytical thought in card games: world class bridge and poker players. In contrast to Palacios-Huerta and Volij's soccer players, we find little evidence that real-world experience transfers to the lab for either subject pool. These poker players, who have vast experience with high-stakes randomization, play the mixed strategy equilibrium in these games roughly as often as our student subjects. We proceed to test whether a sample of professional soccer players from the U.S. play according to theory. Again, we observe little evidence in line with the predicted mixed strategy equilibrium. Coupling complementary experimental treatments that pit professionals against preprogrammed computers with answers to post-survey questionnaires permits us to explore why professionals do not perform well in the lab.

JEL: C72 (Noncooperative games); C9 (Design of Experiments)

Key words: Mixed Strategy, Minimax; Laboratory Experiments

I. Introduction

John von Neumann's (1928) Minimax theorem preceded Nash Equilibrium as the first general framework for understanding of play in strategic situations.[1] The underlying logic of the minimax argument has subsequently been applied broadly—from models of firm, animal, and plant competition to the optimal actions of nations at war. In zero-sum games with unique mixed-strategy equilibria, minimax logic arguably has more intuitive appeal than the standard Nash Equilibrium argument: one needs to randomize strategies in order to prevent exploitation by one's opponent. A nagging issue, unfortunately, is that subjects in laboratory studies typically do not play minimax (see, e.g., Lieberman, 1960; 1962; Brayer, 1964; Messick, 1967; Fox, 1972; Brown and Rosenthal, 1990; Rosenthal et al., 2003). O'Neill (1991, p. 506) aptly summarizes these results by noting that "by the mid-1960s, non-cooperative theory had received so little support that laboratory tests ceased almost completely." Perhaps this finding should not have come as a surprise given that experimental subjects are unable to produce a random series of responses, even when directed to do so (Rapaport and Budescu, 1992).

Recent evidence from field data has provided a renewed sense of optimism, however. Walker and Wooders (2001) analyze serve choices in Grand Slam tennis matches and report similar win rates across various strategies, a result consistent with the minimax equilibrium prediction. Yet, they find that the players switch from one strategy to another too often, a result at odds with minimax theory but consonant with laboratory experimental evidence. Hsu et al. (2007) extend Walker and Wooders (2001) by analyzing a broader tennis data set and report even stronger evidence of equilibrium play: not only

---

[1] There is a debate about whether von Neumann was actually the originator of such thought. Some argue that the great French mathematician Emile Borel preceded von Neumann, but found the idea not worth pursuing because he viewed it as likely false.

are win rates similar across strategies, but individual play is serially independent, as predicted by minimax theory. Complementing these data are results from Chiappori et al. (2002) and Palacios-Huerta (2003), who examine penalty kicks in professional soccer games. Both studies report evidence consonant with von Neumann's minimax theorem— winning probabilities are identical across strategies and choices are serially independent.

Combined, these two strands of literature present an important puzzle: why do controlled laboratory tests of the minimax theorem systematically fail while less controlled tests using field data appear to succeed? Perhaps the tests using field data lack statistical power to reject minimax play. Or, maybe the laboratory has not provided the appropriate environment—for example, crucial experience and context—for subjects to learn the gaming rules to produce equilibrium play.

Palacios-Huerta and Volij (2007) provide an important resolution to this puzzle. Using two standard zero-sum laboratory games with both amateur and professional soccer players as experimental subjects, they report striking evidence that both subject pools use minimax strategies in laboratory experiments. To the best of our knowledge, this study is seminal in that it is the first laboratory game that shows experimental subjects can both i) play strategies in the predicted equilibrium proportions, and ii) generate a sequence of choices that are serially independent. The authors attribute their result to an ability of the soccer players to transfer expertise they have learned shooting penalty kicks to laboratory card games.

We suspect that if this result had surfaced earlier the past several decades of experimentation on noncooperative games would have proceeded in a much different direction. A result as surprising and important as this one warrants further exploration,

including extensions to other domains of individual expertise and across subjects who naturally use equilibrium mixing as well as those who do not. The analysis should also include games where players not only compete against other players, but also compete against computers preprogrammed to play either an optimal strategy or to play exploitable strategies. This type of examination permits control over the strategies of a subject's opponents, allowing an exploration of the underpinnings of why theory passes or fails.

This paper begins by reporting data from three distinct subject pools—undergraduate students, professional bridge players, and professional poker players—playing the same two zero-sum laboratory games as in Palacios-Huerta and Volij (2007). Members of both of our professional subject pools have extensive experience thinking analytically in card games. For our purposes, however, there is one crucial difference between bridge and poker players: there is virtually no role for mixed strategies in bridge. In contrast, randomization is an integral component of skillful poker, as Friedman (1971, B-764) notes: "it is quite clear to those who have played much poker that some sort of mixed strategy…must be used." Professionals clearly recognize the need for randomization as well. A world champion poker player writes of the importance of what he calls a "balanced" (i.e. mixed) strategy: "…you will have to learn to vary both your raises and calls, as well as the actual size of your bets, to avoid giving your opponents a read on your style. (Harrington and Robertie, 2005, p. 52)"[2]

The results we obtain using college students as experimental participants parallel those previously reported in the literature: student play is shown to deviate from the

---

[2] Interestingly, Harrington and Robertie (2005) also recognize the difficulty of following this strategy without the aid of a randomizing device (p. 53): "How do you implement a balanced strategy? It's hard to remember exactly what you did the last four or five times a given situation appeared, but fortunately you don't have to. Just use the little random number generator that you carry around with you all day. What's that? You didn't know you had one? It's the second hand on your watch."

theoretical predictions, as has been uniformly observed in past research (see, e.g., Lieberman, 1960; 1962; Brayer, 1964; Messick, 1967; Fox, 1972; Brown and Rosenthal, 1990; Rosenthal et al., 2003; Palacios-Huerta and Volij, 2007). Unlike Palacios-Huerta and Volij (2007), however, play from subjects drawn from both of our professional participant pools is statistically indistinguishable from student play, and far from minimax: professionals do not play strategies in the predicted proportions, and their choices are serially dependent. This finding holds when the professionals compete against other players as well as when they are informed that they are playing against a computer preprogrammed to exploit individual deviations from optimal play.

These results induced us to test our own sample of professional soccer players drawn from three Major League Soccer (MLS) teams: the Los Angeles Galaxy, Chivas USA, and Real Salt Lake. Given that we had limited time and opportunity to run the experiment, we chose to play only one of the zero-sum games—the 4x4 matrix game—which produced the most compelling results presented in Palacios-Huerta and Volij (2007). Similar to the results from our other treatments, professional soccer players do not mix in the predicted proportions, and their choices are serially dependent.

In light of the large body of psychological evidence that reports limited transfer of learning across tasks (Loewenstein, 1999), we suspect that our failure to find minimax play in the lab is due to the fact that the two zero-sum games themselves are not ideal representations of what the subjects actually face in the field, or at least the players are not recognizing them as such. To dig a level deeper into this hypothesis, we conducted a post-experimental survey inquiring how the professional soccer players interpreted the experimental game in which they had just participated. Consonant with the data patterns

observed, not one soccer player who participated in the experiment spontaneously responded that the experiment reminded them of penalty kicks. Even when specifically prompted with a question about penalty kicks, many of the subjects saw little connection between the lab game and penalty kicks.

The remainder of our study proceeds as follows. Section II presents the experimental design. Section III summarizes our results. Section IV concludes.

II. Experimental Design

We chose to follow the recent literature (including Palacios-Huerta and Volij, 2007) and use two different matrix games in our tests, which we described to subjects as "Hide and Seek" and "Four-Card Barry," respectively. Hide and Seek is a 2x2 matrix game taken from Rosenthal et al. (2003); Four-Card Barry is a 4x4 matrix game developed by O'Neill (1987).[3] As Figures 1a and 1b demonstrate, both games are two player zero-sum variants, with simple enough strategy spaces that one would reasonably expect the minimax hypothesis to have predictive power

A. Student Subjects

To provide a baseline of comparison, our examination begins with an exploration of undergraduate student play in these two games. We included a total of forty-six students from the University of Arizona—twenty-two students participated as subjects in two sessions of Hide and Seek and twenty-four participated as subjects in three sessions of Four-Card Barry. No subject participated in more than one session.

---

[3] The O'Neill experiment was seminal in the sense that it moved experimental tests of minimax theory to an environment that required fewer assumptions on players' utility functions. Prior to O'Neill (1987), previous experimenters assumed that utility depended only on the players' own payoff and furthermore was a linear function of that payoff. By restricting the game to two outcomes—win or lose the same dollar amount— O'Neill was able to construct a matrix with the property that a players' minimax strategy is invariant over reasonable utility functions. Mark Walker and John Wooders suggested the name "Four-Card Barry" to us.

Following Rosenthal et al. (2003), in the Hide and Seek treatment we had each pair of participants sit opposite each other, with a game conductor sitting at the side of the table between them. The conductor gave each participant the instructions for Hide and Seek (see Appendix A.1) and read them aloud. Then the participants played several practice rounds until each was sure he or she was ready to play for real money. Participants made their decisions by playing either a red card or a black card,[4] with both players' decisions revealed simultaneously. As specified in the instructions, the game conductor sometimes rolled a six-sided die to determine the winner of a round. Participants played a total of 150 rounds, switching roles after the first 75 rounds. Each round produced a winner who received a payoff; the per-round payoff was $0.25. Each session of Hide and Seek lasted less than an hour.

The payoff matrix in our version of Hide and Seek differs slightly from the 2x2 penalty-kick game studied by Palacios-Huerta and Volij (2007). Those authors created a payoff matrix based on the empirical success percentages from professional penalty kicks. Since we intended to play our game with multiple subject pools, we decided that the soccer success percentages were unimportant. Our primary goal was to implement a game that was easy for the subjects to understand, but was asymmetric and had equilibrium mixing proportions different from 50:50 for each player, in order to reproduce the most important features of the penalty-kick game. For these purposes, we chose the design of Rosenthal et al. (2003) because it can be implemented easily with a six-sided die.

This change led to slightly different equilibrium mixing proportions than in the game used by Palacios-Huerta and Volij (2007). While the penalty-kick game of Palacios-Huerta and Volij (2007) had equilibrium mixing proportions of approximately 36:64 for

---

[4] We used regular playing cards, typically with a ten of a red suit and a ten of a black suit for each player.

Player 1 and 55:45 for Player 2, Hide and Seek has equilibrium mixing proportions of 67:33 for each player. If anything, this change provides theory with a better chance to succeed, as a two-thirds mixing proportion might be cognitively easier to execute than a more complicated proportion like 36:64 or 55:45.

As in Hide and Seek, we implemented Four-Card Barry in a manner that closely followed the literature. Each pair of participants for Four-Card Barry sat opposite each other, with a conductor sitting at the side of the table between them. The conductor gave each participant the instructions for Four-Card Barry (see Appendix A.2) and read them aloud. The participants then played as many practice rounds as they wished until both were ready to play for real money. Participants made their decisions by playing one of their four cards, with both players' decisions revealed simultaneously. Participants played a total of 150 rounds, switching roles after the first 75 rounds. Each round produced a winner who received a payoff of $0.25. Each session of Four-Card Barry typically lasted one hour.

This variant of Four-Card Barry is identical to Palacios-Huertas and Volij (2007), with two minor exceptions. First, since we intended to play this game with professional card players, we chose to use regular playing cards with all four of our subject pools. That is, instead of the colored cards (Red, Brown, Purple, Green) used by Palacios-Huerta and Volij (2007), we gave each player one card of each suit (Spade, Heart, Club, Diamond).[5] Second, to provide additional insights into the ability of subjects to transfer knowledge across tasks, we had the players switch roles halfway through the 150 rounds of the game. Given that previous results show little evidence that play changes over periods (Rosenthal et al., 2003, and Palacios-Huertas and Volij, 2007), this change is also likely innocuous.

---

[5] Each player received all four cards of the same rank, either four nines or four tens. We deliberately avoided using aces in order to avoid having the ace be focal, as most decks of playing cards make the ace of spades much larger than the aces of the other suits.

B. Professional Subjects

Our first departure from the baseline student treatments is to examine play amongst two distinct groups of professionals: expert bridge and poker players. Members of each of these subject pools commonly use analytic thought in card games. Importantly, one major difference between professional bridge and professional poker players is that there is virtually no role for mixed strategies in bridge.[6] In contrast, randomization is an integral component of skillful poker. In a book on the mathematics of poker, Chen and Ankeman (2006, p. 373) reinforce the earlier advice of Harrington, noting that "…balance is the core of what we might call near-optimal play…. A player who plays with perfect balance is unreadable from a strategy standpoint." Negreanu (2007, p. 20), a three-time World Series of Poker champion and 2004 Player of the Year, ranks "being too predictable" as the second most important reason that players lose at poker. Good play in poker calls for frequent reliance on mixed strategies, for exactly the reason specified by minimax theory: to prevent exploitation by one's opponents. Even a conservative estimate of the number of mixed-strategy decisions that a poker player makes each hour is larger than the number of penalty kicks the typical soccer player takes in a year.

Our bridge treatments were conducted in 2006 at the North American Bridge Championships, which took place at the Hyatt Regency Hotel in downtown Chicago. We included a total of 36 participants—14 as subjects in Hide and Seek and 22 as subjects in Four-Card Barry. We followed protocol identical to that discussed above, with a few minor changes. First, recruiting was done by distributing flyers and face-to-face

---

[6] Indeed, until recently, it was thought that there was no role for mixed strategies in bridge. Rubens (2005) and Thompson (2007) have shown that while optimal play in bridge virtually always involves a pure strategy, in very specific situations mixed strategies do perform slightly better.

solicitation onsite. Second, these treatments took place in a hotel conference room. Third, for each of the games the subjects were paid $1 per successful play.[7]

Our poker treatments were conducted in 2006 at the World Series of Poker, in Las Vegas, NV. We included 130 participants: 44 playing Hide and Seek, 52 playing Four-Card Barry, and 34 who played against a preprogrammed computer. Among our sample of players, the self-reported average number of hours spent per week playing poker is 25. Over 87 percent of the players reported to have made money playing poker in the previous year, with average annual earnings of $120,111. The sample included 11 individuals who had won either a World Series of Poker Bracelet or a World Poker Tour event.

With the poker subject pool, we complemented the human-human treatments with two separate computer treatments. We traced such treatments to Messick (1967), who conducted a three choice, two-player repeated experiment where human subjects played against computer algorithms. Messick had human subjects play against three different algorithms: a minimax strategy, an unlimited memory algorithm (see Fudenberg and Levine, 1998), and an algorithm with a limited memory of five-periods. Importantly, the play algorithms were exploitable due to serially correlated play. Fox (1972), Coricelli (2004), Shachat and Swarthout (2004), and Spiliopoulos (2007) serve to extend this line of research in important and interesting directions.

In the spirit of this literature, we employ both an "optimal" and an exploitable algorithm.[8] Our "optimal" treatment has the computer preprogrammed to play minimax for

---

[7] In this regard, we followed Fehr and List (2004) and Haigh and List (2005) in using larger payoffs for the professionals than the students. This was done to provide more comparable payoffs on an opportunity-cost scale and maintain the professionals' attention during the games.

[8] We also attempted to run a third treatment in which the computer was programmed to play a negative serially correlated strategy, but a programming error led that program for Hide and Seek to play the wrong mix of actions as well – a mistake that was only discovered after costing us a great deal of money in large

the first 15 periods. The computer program was also equipped with a simple learning algorithm designed to exploit sub-optimal play on the part of the human subject, which it did for periods 16 onward.[9] In the learning program, the weight given to the computer's prediction of what the human would choose based on the previous pattern of responses increased as the amount of data for a particular subject increased.[10] Given the nature of what we desired to learn from this exercise, the instructions to this game (included as Appendix B.1) which were both read aloud to the subjects and displayed on screen, told the player that "…we have programmed the computer to play the theoretically correct strategy in this game. In addition, any deviations that your play has from this correct style of play will be taken advantage of by the computer." See Appendix B for complete instructions.

Our second computer treatment involved programming the computer to play sub-optimally. In particular, we chose a simple algorithm whereby the computer randomly chose between the available actions with equal probabilities (whereas the optimal mix was 67:33 in the 2x2 game, and 40:20:20:20 in the 4x4 game). This was a static strategy, and no computer learning component was built into this treatment. The only difference

---

payoffs to the humans who quite successfully exploited the computer's bad play! These results are available upon request.

[9] It is important to build in this learning component because if the computer played Minimax equilibrium probabilities regardless of the response of the human opponent, the human should be indifferent between the choice of actions. By programming the computer to exploit play that is off the equilibrium path, we provided subjects with an incentive to properly mix.

[10] To be more specific, this involved a two step procedure: first, we estimated a predicted choice for the human player via a multinomial logit regression model, and calculated a best response to that predicted choice. Second, we averaged that best response together with the theoretical equilibrium ratios (1/3, 2/3 in Hide and Seek; 2/5, 1/5, 1/5, and 1/5 in Four-Card Barry) to provide the computer's mixed strategy for the next round. For example, suppose that in Hide and Seek the logit model predicts that a human Pursuer's next move will be Red. Then Black would be the computer Evader's best response, so we take the simple average of (0% Red, 100% Black) with (2/3 Red, 1/3 Black) to get a mixture of (1/3 Red, 2/3 Black) as the mixture over which the computer randomizes its next play.

In pilot tests we learned to vary the weights over time in order to produce better performance for the computer algorithm. In particular, in Hide and Seek we used the simple average of the predicted logit best-response strategy and the theoretical equilibrium strategy for periods 16-35; in periods after 35 we used a weight of ¾ on our predicted best response and ¼ on the equilibrium play. Four Card Barry was identical except that we used a ¼ weight on our predicted best response and ¾ weight on the equilibrium ratio for periods 16-25.

between the instructions given in the first computer treatment described above and in this treatment was that the sentence saying that the computer had been programmed to play optimally was omitted in this treatment.

In total, we had 34 participants in the computer treatments. To maximize sample sizes, we allocated the participants so that each would play both the Hide and Seek and the Four-Card Barry. But, we did not vary the computer algorithm condition: if a subject was randomly placed in the "optimal" condition, for example, then she was in that condition for both games. And, to control for order effects, the computer program randomly decided whether Hide and Seek or Four-Card Barry would be played first. In aggregate, 21 subjects played both Hide and Seek or Four-Card Barry against an optimally programmed computer, and 13 played both games against a computer programmed to play an exploitable strategy.

In both the human-human treatments and the human-computer treatments, we followed identical protocol to that discussed above, with a few minor changes. As with the bridge players, recruiting was done through the distribution of flyers and face-to-face solicitation at the World Series of Poker venue (the *Rio Hotel*). All treatments were carried out in our hotel suites at the *Rio Hotel*, which hosted the World Series of Poker. Similar to the bridge sessions, subjects were paid $1 per successful play, whether competing against another human or the computer. Finally, the experiment lasted, in most cases, no longer than one hour.

Our third professional subject pool closely follows Palacios-Huertas and Volij (2007), in that we examine play amongst professional soccer players. We obtained permission to run experiments with three MLS teams: the Los Angeles Galaxy, Chivas

USA, and Real Salt Lake. Each of these clubs granted us access to their locker room for only sixty to ninety minutes. Given the time constraint, we limited our soccer player treatments to the 4X4 O'Neill game, which yielded the sharpest results reported in Palacios-Huertas and Volij (2007). We played Four-Card Barry with a total of thirty-two players from the three MLS teams, typically with four or five game conductors simultaneously administering the game to different pairs. These thirty-two players included thirty roster players, plus one team trainer (who had previously played intercollegiate soccer) and one youth player (a goalkeeper) who trained with the team, but was not yet on the official roster. Five of the thirty-two players were goalkeepers, and following Palacios-Huertas and Volij (2007) we made sure to have all five goalkeepers play against non-goalkeepers. Again, we followed identical protocol to that discussed above with bridge and poker players, except that the treatments were carried out in the locker room of the professional soccer clubs.

III. Empirical Results

As noted by Palacios-Huertas and Volij (2007), if subjects are playing the unique minimax equilibrium, then the data generated should conform to four key predictions:[11]

1) For all players combined, the aggregate marginal and joint distributions of actions should correspond to that predicted by equilibrium play,

2) For each particular pair of players, the marginal and joint distribution of actions should correspond to that predicted by equilibrium play,

---

[11] Minimax also generates other predictions regarding play that are less relevant for the questions at hand. For example, there should be no contemporaneous correlation between the choices made by the two players. Note that many of these minimax predictions depend critically on the fact that the same payoff matrix holds across player pairs and iterations of the game. Chiappori et al. (2002) show which of these Minimax predictions survive aggregation and which do not. These aggregation issues are not a concern in the present context because we control the payoffs.

3) Each player should achieve the same payoff when choosing any of the available strategies, and that payoff should be equal to the equilibrium payoff predicted by theory,

4) Actions should be serially uncorrelated.

In what follows, we report our results parsed by subject pool, proceeding in a manner that parallels the experimental design layout.

A. Students

The first column of Table 1 summarizes the results of empirical tests of these four hypotheses in Hide and Seek. The uppermost panel in Table 1 reports sample sizes and tests for asymmetry of play by role.[12] The lower panels include tests of the relevant hypotheses described above. Readers interested in greater detail regarding these findings are directed to the Tables in the Supplementary Appendix. For purposes of comparison, we also report the soccer player results obtained by Palacios-Huertas and Volij (2007).

Panel I in Table 1 reports p-values for rejecting the null hypothesis that the aggregate frequencies match those of minimax play. The first and second rows show results corresponding to the marginal distributions for those playing pursuer (or seeker) and evader, (or hider).[13] The third row reports results for the joint distribution of play, showing whether combinations of actions (for example, black-black plays) match minimax predictions. For the student subjects, shown in column 1, all three of these hypotheses are rejected at the p < .01 level; the aggregate frequencies are clearly not as predicted by minimax.

---

[12] The results on role asymmetry, although not explicitly mentioned earlier or expanded on here, represent an important test of minimax theory as well. These insights complement the findings in Rosenthal et al. (2003), who report significant asymmetries in that nearly every evader in the Hide and Seek game chose left (Red in our case) more often than his pursuer opponent.

[13] These p-values are obtained from Pearson's Chi-square test for goodness of fit, using one degree of freedom for the test of marginal frequencies and three degrees of freedom for the test of joint frequencies.

Panel II in Table 1 continues to focus on action frequencies, but differs from the top panel in reporting results for individual pairs of players, rather than summarizing the aggregate data. Instead of reporting p-values, in this case we show the fraction of individual players for whom we can reject the null hypothesis that the player's actions match minimax play when acting as the pursuer or evader at the $p < .05$ level.[14] The third row in panel II reports a similar statistic, but for the joint play by each pair. In contrast to panel I, *large* numbers represent a violation of minimax in this part of the table.[15]

Focusing on the student subject data in column 1, we find considerable departures from minimax play. For both pursuers and evaders, more than half of the subjects engage in play that is inconsistent with minimax behavior at the $p < .05$ significance level. We are able to reject the hypothesis that both players are jointly following minimax in more than 90 percent of the pairs. In over one-fourth of the pairs, neither of the players actions are consistent with minimax. By combining these insights with those gained in examining the aggregate play in the Supplementary Appendices, we find that play aggregated across all subjects is near equilibrium frequencies, but there is substantially more heterogeneity observed in choice frequencies across subjects than predicted by theory. This result is not only common to other experiments in this class of games (see, e.g., Rosenthal et al., 2003), but is often found in many other types of experimental data.

Panel III of Table 1 differs from panel II in that it focuses on success rates as opposed to frequencies of actions. Each row of this panel reports, for the relevant group, the percentage of subjects (or pairs of subjects) for whom we can reject at the $p < .05$ level

---

[14]As before, we use a Pearson Chi-square test with one degree of freedom for the marginal frequencies and three degrees of freedom for the joint frequencies.

[15] We also examined the frequency with which *neither* player follows minimax, which represents a stronger test of the theory since if one player is following minimax, both players receive the equilibrium payoffs, regardless of the strategy the other follows. The patterns generally follow those discussed above.

the null hypothesis that equal success rates are achieved when playing the two different actions, and that these success rates match the minimax predictions.[16] For over half of the students, we can reject equality of success rates at the p < .05 level. In 11 out of 22 subject pairs we are able to reject that the payoffs *jointly* correspond to minimax.

Finally, panel IV of Table 1 presents the percentage of players for whom we can reject the null hypothesis of no serial correlation in actions, based on the runs test of Gibbons and Chakraborti (1992).[17] The students fare much better on this test than the other tests, with "only" 11 percent of the players exhibiting play that significantly deviates from the no-serial-correlation null. This percentage is significantly less than one would expect from observing results from the scores of individual-level studies in psychology that lead one to conclude that "producing a random series of responses is difficult, if not impossible task to human [subjects], even when they are explicitly instructed" (Wagenaar, 1971, p. 78). However, it is in line with the intuition that subjects competing in dyadic interactions are more likely to yield serially uncorrelated play than in parallel individual choice settings (Rapaport and Budescu, 1992).

In sum, the results we obtain using undergraduate students as subjects parallel those previously reported in the vast literature: winning probabilities are not identical across

---

[16]This is a Pearson Chi-square test with three degrees of freedom that compares the winning and losing frequencies associated with each play to their minimax predictions. The equilibrium win rates for the pursuer and evader are 0.148 and 0.519 respectively on red, and .074 and .259 respectively on black.

[17]The runs test is based on the following distribution:

$$
f(r \mid n_B^i, n_R^i) = \begin{cases} 2\dbinom{n_B^i - 1}{(r/2)-1}\dbinom{n_R^i - 1}{(r/2)-1} \Big/ \dbinom{n_B^i + n_R^i}{n_B^i} & \text{if } r \text{ is even} \\[2em] \left(\dbinom{n_B^i - 1}{(r-1)/2}\dbinom{n_R^i - 1}{(r-3)/2} + \dbinom{n_B^i - 1}{(r-3)/2}\dbinom{n_R^i - 1}{(r-1)/2}\right) \Big/ \dbinom{n_B^i + n_R^i}{n_B^i} & \text{if } r \text{ is odd} \end{cases}
$$

Where r is the number of runs, and n_B^i and n_R^i are the number of black and red choices.
The serial independence hypothesis will be rejected at the 5 percent level if there are too few or too many runs, that is if $F(r \mid n_B^i, n_R^i) < 0.025$ or if $F(r-1 \mid n_B^i, n_R^i) > 0.975$, where $F(r \mid n_B^i, n_R^i) = \sum_{k=1}^{r} f(k \mid n_B^i, n_R^i)$.

strategies and choices are serially dependent (see, e.g., Brown and Rosenthal, 1990; Rosenthal et al., 2003; Palacios-Huerta and Volij, 2007). Consonant with Rosenthal et al. (2003), these results hold whether we examine early periods of play or later periods, suggesting that play is not converging to equilibrium.[18]

Table 2 presents results from Four-Card Barry. The structure is similar to Table 1, with results presented separately for each set of tests.[19] As was the case with Hide and Seek, empirical results mirror the systematic deviations from minimax play previously noted in the literature. Thus, rather than belabor the point by walking the reader through every panel in the tables, we simply summarize these findings by noting that for each of the four tests described above, student subjects deviate importantly from the minimax predictions. In this case, however, there is even more evidence against students being able to produce serially dependent play (panel IV), but less evidence against minimax play in the aggregate data (panel I). These results all hold whether we examine early or late periods of play, as shown in the Supplementary Appendix.

B. Professionals

Results for the expert subjects are contained in the later columns of Tables 1 and 2 and are summarized in a fashion similar to the student data. For both the Hide and Seek and Four-Card Barry games, we find that for all three pools of expert subjects (professional bridge, poker players, and soccer players), the data are not statistically indistinguishable from our student data. Bridge players somewhat outperform the other subject pools in the

---

[18] See the Supplementary Appendices for the results broken down into the first and second halves of the treatment. Across all of our 2x2 treatments, the results for panels I-III are similar for the two halves of play. In all cases, however, the frequency of rejection for serially correlated play is lower as subjects gain experience with the game.

[19] The Chi-squared tests for the 4x4 games use 3 degrees of freedom for the marginal distributions and 15 degrees of freedom for the joint distributions in panels I and II. In Panels III and IV, play is broken down into diamond versus non-diamond plays and the analyses proceed as in the 2x2 game.

Hide and Seek game, but they pale in comparison to the soccer players in Palacios-Huertas and Volij (2007), whose results we report in the rightmost columns of the tables. For instance, we are able to reject consistency with minimax of the joint frequencies of play for 57 percent of our bridge players; in Palacios-Huertas and Volij(2007), there is not a single rejection for professional soccer players and only one of ten pairs rejects among their college soccer players. Similarly, for bridge players we reject that payoffs are equal to minimax expectations for over one-third of the pairs, compared to no rejections for professional soccer players in Palicios-Huertas and Volij (2007).

In Four-Card Barry, the subject pool that performs the "best" among those we sampled is the professional soccer players, shown in column 4 of Table 2. Once again, however, play of our best subject pool does not rival the results observed for soccer players in Palacios-Huertas and Volij (2007). For example, on the tests of aggregate frequencies in panel I, we reject minimax of our soccer professionals at the $p < .001$ level, whereas in Palacios-Huertas and Volij (2007) the p-value is over .90.[20] Overall, in almost all of our subject pools we strongly reject minimax, a result not in accord with Palacios-Huertas and Volij (2007). One exception is the runs test displayed in panel IV in the tables, though our subjects remain inferior to the soccer players in Palacios-Huertas and Volij (2007).

In sum, we are left to conclude that that none of our subject pools, including professional soccer players, play these laboratory games in accordance with the predictions of minimax, even though there is evidence that they likely employ such strategies in the field (Walker and Wooders, 2001; Chiappori et al., 2002; Palacios-Huerta, 2003;

---

[20] It is worth noting that if we restrict our poker sample to the seven pairs in which both players fit our definition of "world class" (i.e. had won a World Series of Poker bracelet, a World Poker Tour tournament, or reported earning more than $100,000 playing poker in the previous 12 months), play is closer to minimax (especially in panel II, which corresponds to frequencies of play), but does not approach the Spanish soccer players.

Harrington and Robertie, 2005; Chen and Ankeman, 2006; Negreanu, 2007; Hsu et al., 2007). It appears that for our experimental subjects, what happens in the field stays in the field.

C. Understanding why play deviates from minimax

There are several different plausible explanations as to why subjects in our sample fail to play the minimax strategy. A first explanation is that they would like to play minimax, but they are unable to do so because they cannot cognitively solve for the equilibrium, or they are cognitively able but deem the costs too prohibitive. A second, very different explanation is that the players do not believe that their opponents will play minimax. If opponents systematically deviate from minimax (or are expected to deviate), then minimax is no longer the best response because the opponent's strategy is exploitable. For instance, if my opponent is the row player in Four-Card Barry and chooses diamond too frequently, then I also want to choose diamond too frequently. Similarly, if I know that my opponent exhibits negative serial correlation, then that will induce correlation in my choices as well. A third class of explanations lies at the foundation of the experimental environment: the nature and context of the constructed situation did not induce the professionals to retrieve the relevant cognitive tool kit to play optimally. We use two approaches to provide a deeper understanding of our results: computers as opponents and post-experimental surveys.

We focus first on play against preprogrammed computers. As noted above, in the first of these treatments, the computer was programmed to play minimax (and then to exploit its competitor if possible), whereas in the second treatment, the computer program involved continuous sub-optimal play.

Table 3 reports results for the two computer treatments, with the optimally programmed opponent shown in columns 1-2 and the naïve computer opponent in columns 3-4. Results are shown separately for the Hide and Seek and Four-Card Barry games. The top portion of Table 3 presents the same four tests included in the preceding tables, except that we restrict our hypothesis tests to the behavior of the human player.[21]

Importantly for our purposes, even when faced with an opponent programmed to initially play minimax and to only deviate from that strategy in response to non-minimax play by the subject, poker players' actions are not consistent with minimax theory.[22] For a majority of the tests, empirical results are quite similar to the results obtained from the human-human interactions. Indeed, if anything, the runs test reveals that players are more likely to exhibit serially dependent play when competing against the computer, a result consonant with Rapaport and Budescu (1992) if subjects in this treatment interpreted the situation as an individual, non-competitive, choice.[23] By and large, our findings are consistent with the literature that makes use of human-computer interaction (Messick, 1967; Fox, 1972; Coricelli, 2004; Shachat and Swarthout, 2004; Spiliopoulos, 2007).

The bottom panel of Table 3 reports other relevant results for these treatments, including average payoffs as well as the fraction of players who "beat" the computer in the sense of winning more than half of the trials.[24] As expected, the computer programmed to

---

[21] As would be expected, the play of the naively programmed computer is nearly always rejected as being consistent with minimax. Less frequently, but often, the play of the optimally programmed computer is also rejected since it deviates from minimax in response to sub-optimal play on the part of the human subject. These results are presented in the Supplementary Appendix.

[22] When we divide each player's actions into two equal size sets corresponding to the first 75 and last 75 plays, we find similar results (see the Supplementary Appendix).

[23] In Four-Card Barry, the "world class" poker players perform better than the other players in the sample, although the sample size is small. In Hide and Seek, "world class" players do not play better than the others.

[24] Similar to the human-human treatments, the computer and the player switched roles (row vs. column) halfway through the experiment so that each had the same expected value in terms of wins when playing equally well.

play optimally slightly outperforms its human opponents: humans win 49.6 percent of the payouts in the Hide and Seek game and 48.5 percent in Four-Card Barry. Note that the longer the game proceeds, the better the computer does relative to the human subjects. Recall that the computer was programmed to put increasing weight over the course of each trial on choices that exploit systematic deviations from minimax by the human opponent. This approach proved effective: whereas humans won slightly more than 50 percent of the early trials (suggesting that our human subjects were lucky on average since the computer played minimax in the early periods, guaranteeing both players an expectation of 50 percent), by the final third of the treatments the humans won only 48 and 45.5 percent of the payoffs in Hide and Seek and Four-Card Barry. These results are consonant with findings from Messick (1967), who reported that subjects' payoffs were not significantly different from equilibrium payoffs when playing against the minimax algorithm.

Our subjects fared better against the computer programmed to play naïve, non-minimax strategies, particularly in Four-Card Barry, where humans obtained 57.5 percent of the payouts, and 12 of 13 humans earned more than the computer. This result accords with insights gained Messick (1967), Fox (1972), Coricelli (2004), and Shachat and Swarthout (2004), who report that subjects have some propensity in these games to exploit non-optimal play. While our subjects were able to exploit effectively, they did not perform optimally. Given that the naïve computer chose randomly between the four strategies with equal probability, the optimal human strategy is to always play diamond when the row

player and never play diamond when the column player, yielding an expected payout of 62.5 percent. None of our subjects realized this payoff level.[25]

In stark contrast to Four-Card Barry, humans did remarkably poorly against the naïvely programmed computer in Hide and Seek, winning only 50.9 percent of the time, when the pure-strategy best response to the computer's strategy would yield an expected payout of 75 percent. Amazingly, only 8 of 13 humans beat the naïve computer in Hide and Seek. Overall, both sets of results mirror Fox (1972), who finds that subjects adjusted their play in the direction of a best response, but did not play optimally.

Our second approach to learning why minimax theory met with limited success in our lab experiments is to use post-experimental surveys for the soccer treatments to provide insights into whether the constructed experimental environment adequately signaled the nature of the task. Given that our results using U.S. professional soccer players differ from those of Palacios-Huertas and Volij's (2007) soccer players (both professionals and college students with soccer experience), it is important to understand the context our subjects placed on the game. As Harrison and List (2004) note, when one uses neutral context in the experimental instructions, an important risk is that the experimenter loses control because the subject places a different context on the game than the experimenter expects. Indeed, Palacios-Huertas and Volij (2007) attribute the strong performance of soccer players to their prior experience with penalty kicks, an assertion that critically relies on the fact that the experimental subjects place the "correct" context on the game (see also Levitt and List, 2007, on this issue).

---

[25] Interestingly, the humans did as well against the naïve computer early in the trials as when they had more experience with the strategy the computer played. This suggests that relatively little learning took place on the part of the humans, or what learning that did occur happened very quickly.

We used two survey instruments each given to a part of the professional soccer sample. These survey instruments are included in Appendix C. Included in the survey for some of the subjects is the question "Does this game remind you of any other games?" None of the twelve soccer players asked this question spontaneously made the link between the experimental game and penalty kicks. Among the remaining soccer players, when prompted for a comparison between the lab games and penalty kicks, 4 players responded that they saw no comparison at all, 2 said that they only thought about penalty kicks *after* the question was asked, 9 said that they were somewhat comparable, and 5 gave an outright yes. Interestingly, when asked their strategies on penalty kicks, 44 percent of the players reported playing pure strategies (e.g., always kick left), highlighting the fact that very few professional soccer players ever get the chance to take penalty kicks in games and suggesting that few subjects viewed our experimental environment as having a direct parallel to their field of expertise.

IV. Conclusions

Determining the conditions under which people play mixed strategies is a question of fundamental importance in economics. O'Neill (1991, p. 503) remarks that "the most basic idea in game theory is the minimax argument." Judged by past laboratory experimentation, this most basic idea has been met with limited success. An exciting new study due to Palacios-Huertas and Volij (2007) has revived this area of research by generating data that is aligned with some of the main predictions of minimax theory. Using two standard zero-sum games, they report striking evidence that soccer players use minimax strategies in laboratory experiments: their subjects equate their strategies' payoffs to the equilibrium ones, and generate a sequence of choices that are serially independent.

When laboratory results using experienced subjects i) are so starkly different than received results using student subjects, ii) match the predictions of theory quite closely and iii) match how the experts themselves behave in the field, it is likely due to the fact that the experiment is designed in a manner that permits the expert subjects to transfer appropriate heuristics and learned rules of thumb from the field to the laboratory domain. To advance the economic science in such cases, it is important to extend the analysis to other appropriate domains to provide a deeper explanation for the underpinnings of such results.

This paper extends the Palacios-Huertas and Volij (2007) analysis to other domains of individual expertise—across subjects who naturally use equilibrium mixing as well as those who do not—and to games where players not only compete against other players, but also compete against preprogrammed computers. Our evidence is not optimistic for the literal point predictions of the minimax hypothesis. Across three expert subject pools: professional bridge, professional poker, and professional soccer players, we find evidence rejecting the minimax hypothesis. Even within the subset of professionals who have vast experience with high-stakes randomization, world class poker players, the minimax predictions do not fare well.

Although these subjects might be able to randomize effectively in their chosen line of business, they seemingly had difficulty transferring their particular field situation to the specific lab task. Clearly, subjects come to experiments with rules of behavior learned in the outside world. Depending on whether the specific context of the lab game cues the proper rules of thumb, radically divergent results can be obtained. Harrison and List (2007), for instance, examine the behavior of professional bidders in their naturally occurring environments. In their real-world bidding, such subjects do not constantly fall

prey to the winner's curse. When the expert bidders are placed in unfamiliar roles, however, they often fall prey to the winner's curse, just as happens in the lab. Our results combined with their insights underscore an important methodological point: slight changes in context can have profound behavioral effects, whether students or professionals are the experimental participants.[26]

References

Brayer, A.R. 1964. An experimental analysis of some variables of minimax theory. *Behavioral Science* 9 (1): 33–44.

Brown, J.N., and R.W. Rosenthal. 1990. Testing the minimax hypothesis: A re-examination of O'Neill's game experiment. *Econometrica* 58(5): 1065–1081.

Budescu, D.V., and A. Rapoport. 1992. Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology* 121 (3): 352-363.

Chen, B., and J. Ankenman. 2006. *The Mathematics of Poker,* Pittsburgh, PA: ConJelCo.

Chiappori, P.A., S. Levitt, and T. Groseclose. 2002. Testing mixed strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review* 92 (4):1138–1151.

Coricelli, G. 2004. Strategic interaction in iterated zero-sum games. Unpublished Manuscript. http://economics.eller.arizona.edu/downloads/working_papers/coricelli.pdf

Fehr, E., and J. List. 2004. The hidden costs and returns of incentives: Trust and trustworthiness among CEOs. *Journal of the European Economic Association* 2 (5): 743-71.

Fox, J. 1972. The learning of strategies in a simple, two-person zero-sum game without saddlepoint. *Behavioural Science.* 17 (3): 300–308.

---

[26] More generally, the importance of context in games is illustrated in the vast research on the Wason selection task. The Wason selection task is often used to assess what information is necessary in order to test the truth of an abstract logical reasoning problem: if P then Q. A typical experiment will present a rule and ask subjects whether the rule is being violated. Consider the rule: *If a card has a J on one side, then it has a 5 on the other side.* Subjects are aware that on the particular set of cards, each one has a letter on one side and a number on the other side. Four cards are shown, a P, P', Q, and Q' card and subjects are asked to pick correctly the two cards to turn over to verify the rule. Roughly 1 in 5 subjects answer correctly: to turn over the P and Q' cards. When an identical structured problem is given a context: *If Derek hits the golf ball, then it will go in the water,* the success rate skyrockets as most subjects readily realize that they need to know what happened when Derek hit the ball (P) and the action associated with the ball not going in the water (Q').

Friedman, L. 1971. Optimal bluffing strategies in poker. *Management Science*, 17 (12): B764-B771.

Fudenberg, D., and D. Levine. 1998. Learning in Games: Evidence from coordination games. *European Economic Review* 42 (3): 631-639.

Gibbons, J.D., and S. Chakraborti. 2003. *Nonparametric statistical inference, 4th edition.* New York: Marcel Dekker, Inc.

Haigh, M., and J. List. 2005. Do professional traders exhibit myopic loss aversion? An experimental analysis. *Journal of Finance* 60 (1): 523-534.

Harrington, D., and D. Robertie. 2005. *Harrington on Hold'em: Expert strategy for no-limit tournaments, Volume 1: Strategic play,* Henderson, NV: Two Plus Two Publishing.

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

Harrison, G.W., and J. List. 2007. Naturally occurring markets and exogenous laboratory experiments: A case study of the winner's curse. *Economic Journal* (forthcoming)

Hsu, S-H., C-Y. Huang, and C-T. Tang. 2007. Minimax play at Wimbledon: Comment. *The American Economic Review* 97 (1): 517-523.

Levitt, S., and J. List. 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21 (2): 153-174.

Lieberman, B. 1960. Human behavior in a strictly determined 3×3 matrix game. *Behavioral Science* 5(4): 317–322.

Lieberman, B. 1962. "Experimental studies of conflict in some two-person and three-person games." In *Mathematical Models in Small Group Processes*, edited by J.H. Criswell, H. Solomon, and P. Suppes. Stanford, CA: Stanford University Press. 203–220.

Loewenstein, G. 1999. Experimental economics from the vantage-point of behavioral economics." *Economic Journal* 109(453): F23-34

Messick, D.M. 1967. Interdependent decision strategies in zero-sum games: A computer-controlled study. *Behavioral Science* 12 (1): 33–48.

Negreanu, D. 2007. *Hold'em wisdom for all players.* New York, NY: Cardozo.

O'Neill, B. 1987. A non-metric test of the minimax theory of two-person zerosum games. *Proceedings of the National Academy of Sciences* 84 (7): 2106-2109.

O'Neill, B. 1991. Comments on Brown and Rosenthal's reexamination. *Econometrica* 59 (2): 503-507.

Palacios-Huerta, I. 2003. Professionals play minimax. *Review of Economic Studies* 70 (2): 395-415.

Palacios-Huerta, I. and O. Volij. 2007. Experientia docet: Professionals play minimax in laboratory experiments. *Econometrica* (forthcoming).

Rosenthal, R.W., J. Shachat, and M. Walker. 2003. Hide and seek in Arizona. *International Journal of Game Theory* 32 (2): 273-293.

Rubens, J. 2005. Mixing it up. *The Bridge World* 77 (3): 50-54.

Shachat, J., and T.J. Swarthout. 2004. Do we detect and exploit mixed strategy play by opponents? *Mathematical Methods of Operations Research* 59 (3): 359-373.

Schelling, T.C. 1960. *The strategy of conflict.* Cambridge, MA: Harvard University Press.

Spiliopoulos, L. 2007. Do repeated game players detect patterns in opponents? Revisiting the Nyarko and Schotter belief elicitation experiment. *Munich Personal RePEc Archive* Paper No. 2179. http://mpra.ub.uni-muenchen.de/2179

Thompson, M. 2007. On queen-jack-low opposite ace-fifth. *The Bridge World* (online edition only). http://www.bridgeworld.com/default.asp?d=esoterica&f=mark.html

Von Neumann, J. 1928. Zür theorie der Gesellschaftsspiele. *Mathematische Annalen* 100 (1): 295–320.

Walker, M., and J. Wooders. 2001. Minimax play at Wimbledon. *American Economic Review* 91 (5): 1521–38.

Wagenaar, W.A. 1971. Serial non-randomness as a function of duration and monotony of a randomization task. *Acta Psychologica* 35 (1): 78–87.

Figure 1a: Payoff Matrix for the 2x2 Game

		Evader	
		Black	**Red**
Pursuer	**Black**	Die Roll: 1 or 2 (0, 1)	No die roll (0, 1)
	Red	No die roll (0, 1)	Die Roll: 1 or 2 (1, 0)

Figure 1a shows the payoff matrix for the 2x2 game. The pursuer's payoff is given first, followed by the evader's payoff. If both cards match, a die roll determines who wins the round. Cells 1 and 4 show the payoffs for low values of the die. High values yield the opposite payoff.

Figure 1b: Payoff Matrix for the 4x4 Game

		Column			
		Club	**Heart**	**Spade**	**Diamond**
	Club	(1, 0)	(0, 1)	(0, 1)	(1, 0)
Row	**Heart**	(0, 1)	(1, 0)	(0, 1)	(1, 0)
	Spade	(0, 1)	(0, 1)	(1, 0)	(1, 0)
	Diamond	(1, 0)	(1, 0)	(1, 0)	(0, 1)

Figure 1b shows the payoff matrix for the 4x4 game. The row player's payoff is given first, followed by the column player's payoff. Row wins if two non-diamond cards are played and the suits match, or if a diamond card and a non-diamond card are played. Column wins if two non-diamond cards are played and the suits do not match, or if two diamond cards are played.

Table 1: Summary of Results across Subject Pools in the 2 x 2 Game

Source:	Levitt, List & Reiley			Palacios-Huerta & Volij	
Test:	College Students	Bridge Players	Poker (Overall)	Soccer Pros	Soccer College
#Pairs of Players	11	7	22	20	20
#Pairs of Roles	22	14	44	20	20
t-test for asymmetry in plays by role	0.124	0.645	0.001	N/A	N/A
I. Minimax play at aggregate level					
Chi-square test for minimax play:					
Pursuer	<0.001	<0.001	<0.001	0.072	<0.001
Evader	<0.001	0.001	<0.001	<0.001	<0.001
Joint play	<0.001	<0.001	<0.001	0.007	N/A
II. Minimax play at individual level					
Rejections at 5 percent:					
Pursuer	59.09%	42.86%	68.18%	5.00%	10.00%
Evader	54.55%	35.71%	52.27%	5.00%	5.00%
Joint Play	90.91%	57.14%	75.00%	0.00%	5.00%
Neither Player	27.27%	7.14%	40.91%	0.00%	N/A
III. Equality of Success Rates Across Strategies and to the Equlibrium Rate					
Rejections at 5 percent:	63.64%	35.71%	67.05%	0.00%	10.00%
IV. Runs Tests					
Rejections at 5 percent	11.36%	7.14%	9.09%	2.50%	3.75%

Table 1 reports results for the 2x2 matrix game based on the game used by Rosenthal et al (2003). The Columns correspond to the different subject pools tested, while the rows report results for each test. The last two columns report results for a similar experiment carried out by Palacios-Huerta and Volij (2007). Panel I shows p-values from Pearson's Chi-square test for goodness of fit of aggregate frequencies to minimax predictions. P-values for the marginal frequencies of the pursuer and evader are shown in the first two rows, while the third row shows p-values for combinations of plays by both players. The test uses one degree of freedom for the marginal distribution of play and three for the joint distribution. Panel II shows the percentage of individuals (or pairs) that we reject at the 5% level for this same Chi-square test. Panel III shows the percentage of individuals for whom we reject the null hypothesis that success rates are equal across strategies and equal to the equilibrium success rate.

Table 1 continued

The Pearson test statistic is also used in this case, but with winning and losing probabilities for each play rather than with frequencies. Thus, it is a chi-square test with 3 degrees of freedom. Panel IV presents the percentage of players for whom we can reject the null hypothesis of no serial correlation in actions, based on the runs test of Gibbons and Chakraborti (1982) which has the following distribution:

$$
f(r \mid n_B^i, n_R^i) = \begin{cases} 2\binom{n_B^i - 1}{(r/2)-1}\binom{n_R^i - 1}{(r/2)-1} \Big/ \binom{n_B^i + n_R^i}{n_B^i} & \text{if } r \text{ is even} \\[2ex] \left[\binom{n_B^i - 1}{(r-1)/2}\binom{n_R^i - 1}{(r-3)/2} + \binom{n_B^i - 1}{(r-3)/2}\binom{n_R^i - 1}{(r-1)/2}\right] \Big/ \binom{n_B^i + n_R^i}{n_B^i} & \text{if } r \text{ is odd} \end{cases}
$$

Where r is the number of runs, and n_B^i and n_R^i are the number of black and red choices.

The serial independence hypothesis will be rejected at the 5 percent level if there are too few or too many runs, that is if $F(r \mid n_B^i, n_R^i) < 0.025$ or if $F(r-1 \mid n_B^i, n_R^i) > 0.975$, where

$$F(r \mid n_B^i, n_R^i) = \sum_{k=1}^{r} f(k \mid n_B^i, n_R^i).$$

Table 2: Summary of Results across Subject Pools in the 4 x 4 Game

Source:	Levitt, List & Reiley				Palacios-Huerta & Volij	
Test	College Students	Bridge Players	Poker (Overall)	Soccer Pros	Soccer Pros	Soccer College
# Pairs of People	12	11	26	16	20	20
# Pairs of Roles	24	22	52	32	20	20
t-test for asymmetry of play by role	0.188	0.074	0.654	0.376	N/A	N/A
I. Minimax play at aggregate level						
Chi-square test for minimax play:						
Row Player	0.320	0.362	0.253	0.001	0.956	0.890
Column Player	0.008	0.006	0.001	<0.001	0.932	0.964
Joint play	0.105	0.030	0.008	<0.001	0.910	N/A
II. Minimax Play at individual level						
Rejections at 5 percent:						
Row Player	33.33%	36.36%	26.92%	28.13%	5.00%	10.00%
Column Player	45.83%	36.36%	34.62%	15.63%	5.00%	10.00%
Joint Play	37.50%	45.45%	30.77%	28.13%	10.00%	5.00%
III. Equality of Success Rates Across Strategies and to the Equlibrium Rate						
Rejections at 5 percent	39.58%	45.45%	35.58%	31.25%	7.50%	7.50%
IV. Runs Tests						
Rejections at 5 percent	27.08%	10.23%	9.13%	3.91%	2.50%	3.75%

Table 2 reports results for the 4x4 matrix game based on the game developed by O'Neill (1987). The Columns correspond to the different subject pools tested, while the rows report results for each test. The last two columns report results for a similar experiment carried out by Palacios-Huerta and Volij (2007). Panel I shows p-values from Pearson's Chi-square test for goodness of fit of aggregate frequencies to minimax predictions. P-values for the marginal frequencies of the row and column players are shown in the first two rows, while the third row shows p-values for combinations of plays by both players. The test uses three degrees of freedom for the marginal distribution of play and fifteen for the joint distribution. Panel II shows the percentage of individuals (or pairs) that we reject at the 5% level for this same Chi-square test. For Panels III and IV, play is divided into two – diamond plays and non-diamond plays – before being analyzed. Panel III shows the percentage of individuals for whom we reject the null hypothesis that success rates are equal across strategies and equal to the equilibrium success rate. The Pearson test statistic is also used in this case, but with winning and losing probabilities for each strategy rather than with frequencies. Thus, it is a chi-square test with 3 degrees of freedom. Panel IV presents the percentage of players for whom we can reject the null hypothesis of no serial correlation in actions, based on the runs test of Gibbons and

Table 2 continued

Chakraborti (1982) which has the following distribution:

$$f(r\mid n_B^i,n_R^i)=\begin{cases} 2\dbinom{n_B^i-1}{(r/2)-1}\dbinom{n_R^i-1}{(r/2)-1}\bigg/\dbinom{n_B^i+n_R^i}{n_B^i} & \text{if } r \text{ is even} \\[2em] \left[\dbinom{n_B^i-1}{(r-1)/2}\dbinom{n_R^i-1}{(r-3)/2}+\dbinom{n_B^i-1}{(r-3)/2}\dbinom{n_R^i-1}{(r-1)/2}\right]\bigg/\dbinom{n_B^i+n_R^i}{n_B^i} & \text{if } r \text{ is odd} \end{cases}$$

Where r is the number of runs, and n_B^i and n_R^i are the number of black and red choices.

The serial independence hypothesis will be rejected at the 5 percent level if there are too few or too many runs, that is if $F(r\mid n_B^i,n_R^i)<0.025$ or if $F(r-1\mid n_B^i,n_R^i)>0.975$, where $F(r\mid n_B^i,n_R^i)=\sum_{k=1}^{r}f(k\mid n_B^i,n_R^i)$.

Table 3: Summary of Results for Subjects Playing against Computers Programmed for Optimal or Naïve Play

Source:	Computer Programmed for Optimal Play		Computer Programmed for Naïve Play	
Test:	2 x 2	4 x 4	2 x 2	4 x 4
Type of Player:	All Players	All Players	All Players	All Players
# Players	21	21	13	13
# Player-Roles	42	42	26	26
I. Minimax Play at Aggregate Level				
Chi-square test for minimax play:				
Evader/Row Player	<0.001	0.132	1.000	<0.001
Pursuer/Column Player	<0.001	<0.001	<0.001	<0.001
II. Minimax Play at Individual Level				
Rejections at 5 percent:				
Evader/Row Player	52.38%	33.33%	76.92%	92.31%
Pursuer/Column Player	57.14%	47.62%	84.62%	100.00%
III. Equality of Success Rates Across Strategies and to the Equlibrium Rate				
Rejections at 5 percent				
Evader/Row Player	38.10%	19.05%	84.62%	100.00%
Pursuer/Column Player	61.90%	42.86%	76.92%	100.00%
IV. Runs Tests				
Rejections at 5 percent	19.05%	15.48%	30.77%	13.46%
V. Mean Player Payoff Relative to Total Payoff				
Overall	0.496	0.485	0.509	0.575
1st 25 Rounds	0.506	0.530	0.512	0.574
2nd 25 Rounds	0.503	0.470	0.500	0.598
3rd 25 Rounds	0.480	0.455	0.514	0.554
Fraction of Players who beat the computer:	4/7	3/7	8/13	12/13

Table 3 reports results for the computer-based experiments. The first two columns correspond to games played on the computer programmed for optimal play, while the last two columns correspond to games played on the computer programmed for naïve play. Panel I shows p-values from Pearson's Chi-square test for goodness of fit of the human player's aggregate frequencies to minimax predictions. P-values for the marginal

Table 3 continued

frequencies of the human player as evader (or row) and pursuer (or column) are shown in the first and second rows. The test uses one (three) degree(s) of freedom for the marginal distribution of play and three (fifteen) for the joint distribution for the 2x2 (4x4 game). Panel II shows the percentage of humans that we reject at the 5% level for this same Chi-square test. For Panels III and IV, play in the 4x4 game is divided into two – diamond plays and non-diamond plays – and then analyzed as in the 2x2 game. Panel III shows the percentage of individuals for whom we reject the null hypothesis that success rates are equal across strategies and equal to the equilibrium success rate. The Pearson test statistic is also used in this case, but with winning and losing probabilities for each play rather than with frequencies. Thus, it is a chi-square test with 3 degrees of freedom. Panel IV presents the percentage of players for whom we can reject the null hypothesis of no serial correlation in actions, based on the runs test of Gibbons and Chakraborti (1982) which has the following distribution:

$$
f(r \mid n_B^i, n_R^i) = \begin{cases} 2 \dbinom{n_B^i - 1}{(r/2) - 1} \dbinom{n_R^i - 1}{(r/2) - 1} \Big/ \dbinom{n_B^i + n_R^i}{n_B^i} & \text{if } r \text{ is even} \\[2ex] \left(\dbinom{n_B^i - 1}{(r-1)/2} \dbinom{n_R^i - 1}{(r-3)/2} + \dbinom{n_B^i - 1}{(r-3)/2} \dbinom{n_R^i - 1}{(r-1)/2} \right) \Big/ \dbinom{n_B^i + n_R^i}{n_B^i} & \text{if } r \text{ is odd} \end{cases}
$$

Where r is the number of runs, and n_B^i and n_R^i are the number of black and red choices. The serial independence hypothesis will be rejected at the 5 percent level if there are too few or too many runs, that is if $F(r \mid n_B^i, n_R^i) < 0.025$ or if $F(r-1 \mid n_B^i, n_R^i) > 0.975$, where

$$
F(r \mid n_B^i, n_R^i) = \sum_{k=1}^{r} f(k \mid n_B^i, n_R^i).
$$

Panel V gives the average player payoff relative to the maximum potential payoff. In equilibrium, the expected payoff is 50 percent.

Appendix A1. Instructions for Hide and Seek

You and the person next to you will be playing a repeated game of pursuit and evasion, or "hide and seek". If you are sitting at the left-hand side of your table, you are the Pursuer; if you are on the right-hand side, you are the Evader. Each of you should have two cards in front of you—a black card and a red card.

The game begins with each player privately choosing either their black or red card, and placing it face down on the table.

Next, both players will turn over their cards. If the cards do not match (i.e., one is red, the other black), then the Evader has succeeded: the Pursuer has not found him, so the Evader wins the hand. Alternatively, if the cards DO match (i.e., both are red or both are black), then the Pursuer has FOUND the Evader, but whether he actually wins the hand depends upon the die number rolled.

In this case, the monitor will randomly select a number from 1 to 6 by rolling this six-sided die. Note that each of the six numbers is equally likely to occur.

The result is determined as follows:

	If the die is 1 or 2:	If the die is 3,4,5, or 6:
If both played red:	Pursuer wins	Evader wins
If both played black:	Evader wins	Pursuer wins
If one red, one black:	No die roll, Evader wins	

Once the winner is determined, the monitor will record who won, the cards played, and the die number rolled. The player's cards will then be returned to them, and a new round will begin.

Play will continue for a number of rounds. Each round that you win will be worth 25¢.

You will play 75 rounds. You will then switch roles and play another 75 rounds.

Each of you has a sheet of paper summarizing the rules.

Appendix A2. Instructions for Four-Card Barry

You and the person next to you will be playing a repeated game of skill called "Four-Card Barry." If you're sitting to the left, you will be called the "Row" player. If you're sitting to the right, you will be called the "Column" player.

You and your opponent compete for chips by each simultaneously playing a card: Heart, Spade, Club, or Diamond. The game begins with each player privately choosing a card, and placing it face down on the table.

Next, both players will turn over their cards.

The winner of each round will be determined as follows:

		Column plays			
		Heart	**Spade**	**Club**	**Diamond**
Row plays	**Heart**	Row	Column	Column	Row
	Spade	Column	Row	Column	Row
	Club	Column	Column	Row	Row
	Diamond	Row	Row	Row	Column

While this might seem difficult to understand, a few examples and practice rounds should make things clearer. Consider the case of both players choosing Heart. As the table denotes, the row player wins in this case.

As you can see, if both players play cards other than diamond, then Row wants the two cards to match suit, and Column wants the two cards not to match suit.   If someone plays a diamond, then Row wins unless *both* players play a diamond, in which case Column wins.

Each round that you win will be worth 25¢. We will play 10 practice rounds before beginning play for real stakes.

You will play 75 rounds. You will then switch roles and play another 75 rounds.

Each of you has a sheet of paper summarizing the rules.

Appendix B1: Instructions for Hide and Seek Versus Computer[27]
Human Subject as Pursuer

You and the computer will be playing a repeated game of pursuit and evasion, or "hide and seek". You are the pursuer, the computer is the evader. You have two cards in front of you on the computer screen, as does the computer—a black 9 and a red 9.

The game begins with both you and the computer choosing either your black or red card; you choose by clicking on the card of your choice on the computer screen.

The computer will then reveal what it played. If the cards do not match (i.e., one is red, the other black), then the Evader (computer) has succeeded: you have not found him, so the Evader wins the hand. Alternatively, if the cards DO match (i.e., both are red or both are black), then you have FOUND the Evader, but whether he actually wins the hand depends upon the die number rolled.

In this case, the computer will randomly select a number from 1 to 6 by rolling a six-sided die. Note that each of the six numbers is equally likely to occur.

The result is determined as follows:

	If the die is 1 or 2:	If the die is 3,4,5, or 6:
If both played red:	Pursuer wins	Evader wins
If both played black:	Evader wins	Pursuer wins
If one red, one black:	No die roll, Evader wins	

Once the winner is determined, the monitor will record who won, the cards played, and the die number rolled. A new round will then begin.

Play will continue for a number of rounds. Each round that you win will be worth $1; each round that you lose will cost you a $1.

Before play begins, we should note we have programmed the computer to play the theoretically correct strategy in this game. In addition, any deviations that your play has from this correct style of play will be taken advantage of by the computer.

---

[27] The instructions for the naively programmed computer are the same as for the optimally programmed computer, except that the last paragraph is omitted.

Appendix B2: Instructions for Hide and Seek Versus Computer
Human Subject as Evader

You and the computer will be playing a repeated game of pursuit and evasion, or "hide and seek". You are the evader, the computer is the pursuer. You have two cards in front of you on the computer screen, as does the computer—a black 9 and a red 9.

The game begins with both you and the computer choosing either your black or red card; you choose by clicking on the card of your choice on the computer screen.

The computer will then reveal what it played. If the cards do not match (i.e., one is red, the other black), then the Evader (you) has succeeded: he has not found you, so the Evader wins the hand. Alternatively, if the cards DO match (i.e., both are red or both are black), then the computer has FOUND the Evader, but whether he actually wins the hand depends upon the die number rolled.

In this case, the computer will randomly select a number from 1 to 6 by rolling a six-sided die. Note that each of the six numbers is equally likely to occur.

The result is determined as follows:

	If the die is 1 or 2:	If the die is 3,4,5, or 6:
If both played red:	Pursuer wins	Evader wins
If both played black:	Evader wins	Pursuer wins
If one red, one black:	No die roll, Evader wins	

Once the winner is determined, the monitor will record who won, the cards played, and the die number rolled. A new round will then begin.

Play will continue for a number of rounds. Each round that you win will be worth $1; each round that you lose will cost you a $1.

Before play begins, we should note we have programmed the computer to play the theoretically correct strategy in this game. In addition, any deviations that your play has from this correct style of play will be taken advantage of by the computer.

Appendix B3: Instructions for Four-Card Barry Versus Computer
Human Subject as Row

You and the computer will be playing a repeated game of skill called "Four-Card Barry." You are the "Row" player.

You and the computer compete for chips by each simultaneously playing a card: Heart, Spade, Club, or Diamond. The winner of each round will be determined as follows:

		Column plays			
		Heart	**Spade**	**Club**	**Diamond**
You play	**Heart**	Row	Column	Column	Row
	Spade	Column	Row	Column	Row
	Club	Column	Column	Row	Row
	Diamond	Row	Row	Row	Column

While this might seem difficult to understand, a few examples and practice rounds should make things clearer. Consider the case of both players choosing Heart. As the matrix denotes, the row player (you) wins in this case.

As you can see, if both players play cards other than diamond, then Row (you) wants the two cards to match suit, and Column (computer) wants the two cards not to match suit. If someone plays a diamond, then Row (you) wins unless *both* players play a diamond, in which case Column (computer) wins.

Each round that you win will be worth $1. We will play 10 practice rounds before beginning play for real stakes.

Before play begins, we should note we have programmed the computer to play the theoretically correct strategy in this game. In addition, any deviations that your play has from this correct style of play will be taken advantage of by the computer.

Appendix B4: Instructions for Four-Card Barry Versus Computer
Human Subject as Column

You and the computer will be playing a repeated game of skill called "Four-Card Barry."
You are the "Column" player.

You and the computer compete for chips by each simultaneously playing a card: Heart,
Spade, Club, or Diamond. The winner of each round will be determined as follows:

		You play			
		Heart	**Spade**	**Club**	**Diamond**
	Heart	Row	Column	Column	Row
Row plays	**Spade**	Column	Row	Column	Row
	Club	Column	Column	Row	Row
	Diamond	Row	Row	Row	Column

While this might seem difficult to understand, a few examples and practice rounds should
make things clearer. Consider the case of both players choosing Heart. As the matrix
denotes, the row player (computer) wins in this case.

As you can see, if both players play cards other than diamond, then Row (computer) wants
the two cards to match suit, and Column (you) wants the two cards not to match suit.   If
someone plays a diamond, then Row (computer) wins unless *both* players play a diamond,
in which case Column (you) wins.

Each round that you win will be worth $1. We will play 10 practice rounds before
beginning play for real stakes.

Before play begins, we should note we have programmed the computer to play the
theoretically correct strategy in this game. In addition, any deviations that your play has
from this correct style of play will be taken advantage of by the computer.

Appendix C1: Soccer Player Survey 1 – Chivas USA and LA Galaxy

Player Survey

This information is completely confidential.

1. What do you think about the game we just played? Do you think your experience in soccer helped you to play this game better?

2. Does this game remind you of penalty kicks?

3. How many different penalty kick strategies do you use? Describe them.

4. Estimate what fraction of the time you use on each of the above strategies.

5. How do you decide where to direct a particular penalty kick?

6. How often do you practice penalty kicks?

Appendix C2: Soccer Player Survey 2 – Real Salt Lake

Player Survey

This information is completely confidential.

1. What do you think about the game we just played?

2. Does this game remind you of any other games?

3. How many different penalty kick strategies do you use? Describe them.

4. Estimate what fraction of the time you use each of the above strategies.

5. How do you decide where to direct a particular penalty kick?

6. How often do you practice penalty kicks?