Experientia Docet: Professionals Play Minimax in Laboratory Experiments*

Ignacio Palacios-Huerta† Oscar Volij‡

January 2006

Abstract

We study how professional players and college students play zero-sum two-person strategic games in a laboratory setting. We first ask professionals to play a 2x2 game that is formally identical to a strategic interaction situation that they face in their natural environment. We find that these subjects play in the laboratory exactly as in the field, that is as the equilibrium of the game dictates: (i) they equate payoffs across strategies, and (ii) generate sequences of choices that are random. In sharp contrast with them, however, we also find that college students play the game far from the equilibrium predictions. We then study the behavior of professional players and college students in the classic O'Neill's 4x4 zero-sum game, a game that none of the subjects have encountered previously, and find the same drastic differences in behavior between these two subject pools. The transfer of skills and experience from the familiar field to the unfamiliar laboratory observed for professional players is relevant, from a methodological perspective, to evaluate the circumstances under which behavior in a laboratory setting may be a reliable indicator of behavior in a naturally occurring setting. From a cognitive perspective, it is useful for research on recognition processes, intuition, and similarity as a basis for inductive reasoning.

---

*We thank Jose Apesteguia, Pedro dal Bó, Vicki Bogan, Juan Carrillo, the late Jean Jacques Laffont, Ana I. Saracho, Jesse Shapiro, and audiences at various seminars and conferences for helpful comments and suggestions. We are especially grateful to the general managers of the soccer clubs that granted access to the players that participated in this study, and to the Universidad del País Vasco for its hospitality. We gratefully acknowledge generous financial support from the Salomon Foundation and the Spanish Ministerio de Ciencia y Tecnología (grant BEC2003-08182). Pedro dal Bó provided the dice. Data and programs are available upon request. Any errors are our own.

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

‡Iowa State University and Ben-Gurion University. Email: oscar@volij.co.il

"We transfer our experience in past instances to objects which are resembling, but are not exactly the same with those concerning which we have had experience. ... Tho' the habit loses somewhat of its force by every difference, yet'tis seldom entirely destroy'd, where any considerable circumstances remain the same."

David Hume, *A Treatise of Human Nature* (1739)

1 Introduction

Over the last two decades experiments in laboratory environments have become an important tool in empirical economic analysis, as insights into behavior that cannot be studied easily in the real world may be obtained in this controlled and artificial setting. Thus, an important question for those areas of economic research that rely on data collected in a laboratory is how applicable are these insights for predicting behavior in natural environments.

This paper addresses this question for situations that involve strategic interaction between subjects. Game theory is, in fact, one of the areas where it is mainly experimental data collected in the laboratory, rather than in natural environments, that are used to inform theoretical developments.[1] One reason for this is that Nature does not always create the circumstances that allow a clear view of the principles at work in strategic situations. Furthermore, naturally occurring phenomena are typically too complex to be empirically tractable.

Laboratory environments provide valuable control of players' information, payoffs, available strategies and other relevant aspects, which is important because game-theoretic predictions are often sensitive to changes in these variables. However, as Harrison and List (2004, pp. 1009-11) remark, "lab experiments in isolation are necessarily limited in relevance for predicting field behavior, unless one wants to insist *a priori* that those aspects of economic behavior under study are perfectly general ... [The reason is that] the very control that defines the experiment may be putting the subject on an artificial margin. Even if behavior on that margin is not different than it would otherwise be without the control, there is the possibility that constraints on one margin may induce effects on behavior on unconstrained margins." These and other doubts about the generalizability and extent to which laboratory results may provide insights into field behavior, demand more elaborate experiments.[2]

[1] Camerer (2003) offers a comprehensive review.

[2] See Weibull (2004) and Lazear, Malmendier and Weber (2005) for other concerns, and Camerer (2003), Harrison and List (2004), and Kagel and Roth (1995) for relevant references on the development of different experiments.

This paper is concerned with one such experiment. In particular, we conduct a conventional experiment with a non-standard subject pool playing a game where players are predicted to choose probabilistic mixtures. Our idea is to take advantage of the opportunity that professional soccer provides in order to develop an "artefactual field experiment" to study a margin not studied previously for games with mixed-strategy equilibria.[3] The suitable circumstances this sport offers are the following: (i) Professional soccer players face a *simple* strategic game that is governed by very detailed rules: a penalty kick; (ii) The formal structure of this game can be reproduced in the laboratory; (iii) Previous research has found that when professional soccer players play this game in the field, their behavior is consistent with the *equilibrium* predictions of the theory. These three distinct characteristics allow us to study whether the skills and heuristics that players may have developed in a familiar field setting transfer to the unfamiliar laboratory, and the extent to which field and laboratory behavior are different. Put it differently, they allow to study, for the first time to the best of our knowledge, the role of laboratory context as a "treatment" in a strategic interaction situation requiring use of mixed strategies. A positive answer to the question of whether field and laboratory behavior are sufficiently similar may then indicate that laboratory findings are reliable for predicting field behavior. A negative answer would suggest the opposite.

We proceed as follows. We first analyze the behavior of professional soccer players in a laboratory setting playing a simultaneous two-person zero-sum 2x2 game that is formally identical to a penalty kick. The equilibrium of the game is unique and requires each player to use a mixed strategy. The procedure we follow makes no references to any type of soccer terminology that may trigger psychological motivations, and we use much lower stakes than in real life. To test our methodological hypothesis, we also implement exactly the *same* controlled laboratory experiment with subjects drawn from the standard subject pool of college students with no soccer experience.

Palacios-Huerta (2003) found that the behavior of professional players in the soccer field was consistent with equilibrium play in every respect: (i) their winning probabilities were statistically identical across strategies; (ii) their choices were serially independent.[4] The results we obtain in this paper can be summarized as follows. We find that professional players continue to behave remarkably consistent with the implications of equilibrium in this entirely different setting. Interestingly, we also find that their behavior is in sharp contrast with that of college students who play quite poorly from the perspective of the equilibrium of the game: they do not to equate winning probabilities across strategies and consistently generate sequences that exhibit negative autocorrelation. We interpret these results as evidence that professionals transfer their learning across these vastly different environments and circumstances.

[3]This term follows the classification suggested in Harrison and List (2004).

[4]See also Chiappori, Levitt and Groseclose (2002) and Azar and Bar-Eli (2005) for further evidence in support of equilibrium behavior.

As such, the nature of the subject pool is important for drawing inferences about the predictive power of the equilibrium of the game.

These results may be of special interest in the context of understanding the determinants of randomization, which is a testable hypothesis shared by every game that admits a mixed strategy equilibrium. An extensive literature in experimental economics, game theory, and psychology consistently finds that subjects are unable to generate i.i.d. sequences in the laboratory. Instead, they tend to exhibit a significant bias against repeating the same choice.[5] We find, however, that professional soccer players do generate perfectly random sequences in the laboratory whereas, consistent with the extensive evidence available in the literature, college students do not.

In an attempt to evaluate whether professional players may behave differently in a game they have not encountered previously in any setting, we ask them to play the 4x4 zero-sum two-person simultaneous game developed in O'Neill (1987), and further studied in Brown and Rosenthal (1990), Shachat (2002) and Walker and Wooders (2001). We again compare their behavior with that of college students. The results show that students behave as previous authors have found, that is far from the predictions of the unique equilibrium of the game. Although we use much greater monetary incentives and subjects play more repetitions than in previous studies of this game, students do not equate winning probabilities across strategies and continue generating sequences of choices that are not random. In sharp contrast with this behavior, we find that professional soccer players play, again, remarkably consistent with equilibrium: (i) their distribution of play is not statistically different from the equilibrium distribution, and (ii) their choices are serially independent. While we have considered various extensions that will be discussed later, such as experiments with students that have soccer experience at the amateur level, these are the main findings of the 2x2 and 4x4 zero-sum games that we would like to emphasize.

We interpret the results that professionals who play a given strategic game in a field setting according to its equilibrium predictions continue to behave as the equilibrium predicts in the laboratory, under lower monetary stakes than in real life, and even when facing an unfamiliar game, as supporting the idea that the vast differences in environments do not undermine the skills these subjects use in the field. The fact that the behavior of professional soccer players is distinctly different from that of college students, the subject pool typically considered in a vast experimental literature, suggests that the game-theoretic *equilibrium* predictions may have greater empirical content than previously considered for explaining behavior in *both* natural and experimental settings. It also suggests that in these games the nature of the subject pool may be a critical ingredient of the laboratory experiment for predicting field behavior.

[5] See Neuringer (2002) and Camerer (1995) for surveys of the relevant literature.

From a methodological viewpoint, we see the artefactual field experiments implemented in this paper as being complementary to traditional laboratory experiments of games where players are predicted to choose probability mixtures.[6]

These results have implications for the literature on cognition and *similarity* as a basis for inductive reasoning.[7] Camerer, Loewenstein and Prelec (2005) review evidence showing how "much of the brain implements 'automatic' processes, which are faster than conscious deliberation" and how "with experience at a task or a problem, the brain seems to gradually shift processing toward brain regions and specialized systems that can solve problems automatically and efficiently with low effort." Similarly, Smith (2005) considers that "human activity is diffused and dominated by unconscious, autonomic, neuropsychological systems that enable people to function effectively without always calling upon the brain's scarcest resource–attention and reasoning circuitry." He also discusses evidence showing how the challenge of any unfamiliar action or problem appears first to trigger a search in the brain to bring to the conscious mind what one knows that is related to the decision context, and how systems built into the brain do their work automatically and largely outside of our conscious awareness.[8]

From this viewpoint, the results in this paper support the hypothesis that cognitive skills may exist beyond those that subjects are aware of, and that these skills are the outcome of learning over an extended period of time in a field setting. The facts that they exist in the context of strategic interaction situations involving mixed strategies, and that they transfer to a highly unfamiliar environment where data to inform economic theories are often obtained, are the main findings of our analysis.

2 Experimental Procedures

We implement two different zero-sum games, each one with two different subject pools: professional soccer players and college students. The experiments were conducted during the period November 2003-October 2004. Each of the two zero-sum games we study was played by 40 professional soccer players working in twenty pairs and 40 college students with no soccer experience working in twenty pairs. We also recruited an additional set of 40 college students with soccer experience at the amateur level working in twenty pairs for each of the two games.

[6]While perfectively competitive games do not represent the entire universe of strategic games involving mixed strategies, a number of authors consider that "zero-sum games are a vital cornerstone" of game theory (Aumann, 1987). See also von Neumann (1928) and Binmore et al (2001).

[7]See, for instance, Hume (1748), Gilboa and Schmeidler (2001), Gigerenzer and Todd (1999), Selten (1998), Simon (1983), and other references therein.

[8]These ideas are also supported in Gneezy, Rustichini and Vostruknutov (2005) who study experimentally how human subjects solve a complex problem, and find that processes involving unconscious reasoning are active in the solution of sophisticated and novel problems.

Next we provide details about the recruiting process for these 240 subjects and other aspects of the experimental procedure, and then describe the experimental designs of the two games we study.

2.1 Subjects

Each subject served in only one game and one session, and those who knew each other were not allowed to participate in the same pair. Sessions lasted about an hour, and subjects received their winnings as payment.

Professional Players. These subjects were recruited from professional soccer clubs in Spain. Professional soccer teams play most of their games in domestic league competitions. As in many other European and South American countries, league competition in Spain is hierarchical. It has three professional divisions: Primera Division with 20 teams, Segunda Division A with 22 teams, and Segunda Division B with 80 teams divided into four groups of twenty teams each.[9] Our subjects come from a number of clubs in the north of Spain, a region with a high density of professional teams. For example, within 150 miles of the city of Bilbao, there are 25 professional soccer clubs participating in league competitions in those three divisions. Teams typically have about 22-26 players in their roster, 2-4 of which are goalkeepers.

Eighty male soccer players (40 kickers and 40 goalkeepers) were recruited from these teams with telephone calls and visiting teams in daily practices.[10] Marca (2005) offers a vitae of every player in Primera Division and Segunda Division A that includes personal information, professional playing history and other records. Forty kicker-goalkeeper pairs were formed randomly using the last two digits of their national ID card with the only requirement that subjects that were currently playing or had played in the past for the same team were not allowed to participate in the same pair.

Undergraduate students. One hundred and sixty male subjects were recruited with fliers around the campus of the Universidad del País Vasco in Bilbao, and by visiting different undergraduate classes. We recruited no subjects majoring in Economics or Mathematics. Half of the subjects we recruited had no soccer experience. The other half had soccer experience at the amateur level as they were required that they should be currently participating in regular league competitions in regional amateur divisions, that is Tercera Division and below. These leagues follow exactly

[9]The next division in the hierarchy, Tercera Division, also includes some players who are professional in that their salaries plus bonuses are similar to the average household salary in Spain. There are 240 teams in Tercera Division in Spain, sorted regionally according to geographical distance into twelve groups of twenty teams each. Teams in divisions lower in the hierarchy, playing in "regional leagues," do not typically have any professional players. Our sample of amateur players comes from Tercera Division and these regional leagues.

[10]No player that had played professionally for less than two years at the time of the experiment was recruited.

the same structure, calendar schedule, and are governed by the same rules (FIFA, 2005) as the professional leagues.

Pairs were formed randomly using the last two digits of their national ID card. For subjects with soccer experience, those that were currently playing or had previously played for the same team were not allowed to participate in the same pair.

2.2 Experimental Designs

2.2.1 Experiment 1: Penalty Kick

Before discussing how the formal structure of a penalty kick may be reproduced in a laboratory setting, it is first useful to go over its basic rules and structure, and the evidence from the field.

In soccer, a penalty kick is awarded against a team which commits one of the ten punishable offenses inside its own penalty area while the ball is in play. The world governing body of this sport, the *Fédération Internationale de Football Association* (FIFA), describes in detail the simple rules that govern this strategic interaction situation in the *Official Laws of the Game* (FIFA, 2005):[11]

- "The ball is placed on the penalty mark in the penalty area.
- The player taking the penalty kick is properly identified.
- The defending goalkeeper remains on the goal line, facing the kicker, between the goalposts, until the ball has been kicked.
- The player taking the penalty kicks the ball forward.
- He does not play the ball a second time until it has touched another player.
- A goal may be scored directly from a penalty kick."

Each penalty kick involves two players: a kicker and a goalkeeper. In the typical kick the ball takes about 0.3 seconds to travel the distance between the penalty mark and the goal line; that is, it takes less than the reaction time plus goalkeeper's movement time to any possible path of the ball.[12] Hence, both kicker and goalkeeper must move simultaneously. The penalty kick has only two possible outcomes: score or no score, actions are observable, and the outcome of the penalty kick is decided almost immediately after players choose their strategies.[13]

[11] The dimensions of the field of play, including the penalty area, the position of the penalty mark, and the distance to the goals, are described in detail in Law 1 of FIFA (2005).

[12] Miller (1998) reports evidence on ball speed, reaction times, and movement times from all the penalty kicks in four World Cups.

[13] The spin of the kick plays no role. There are no second penalties in case a goal is not scored. The initial location of both the ball and the goalkeeper is always the same: the ball is placed on the penalty mark and the goalkeeper positions himself on the goal line, equidistant to the goalposts.

The clarity and simplicity of the rules and structure of this simultaneous one-shot interaction suggest not only that it can be studied empirically, but also that it may be easily reproduced in an artificial setting such as a laboratory. The basic structure of a penalty kick may be represented by the following simple 2×2 game:

$$\begin{array}{c|c|c|} i\backslash j & L & R \\ \hline L & \pi_{LL}, 1-\pi_{LL} & \pi_{LR}, 1-\pi_{LR} \\ \hline R & \pi_{RL}, 1-\pi_{RL} & \pi_{RR}, 1-\pi_{RR} \\ \hline \end{array},$$

where π_{ij} denotes the kicker's probabilities of scoring when he chooses i and the goalkeeper chooses j, for $i, j \in \{L, R\}$. This game has a unique Nash equilibrium when $\pi_{LR} > \pi_{LL} < \pi_{RL}$ and $\pi_{RL} > \pi_{RR} < \pi_{LR}$, which requires each player to use a mixed strategy. When this game is repeated, equilibrium theory yields two sharp testable predictions about the behavior of the players:

1. The probability that a goal will be scored should be the same across strategies for each player, and equal to the equilibrium success probability: p for the kicker and $1-p$ for the goalkeeper, with $p = (\pi_{LR}\pi_{RL} - \pi_{LL}\pi_{RR}) / (\pi_{LR} - \pi_{LL} + \pi_{RL} - \pi_{RR})$.
2. Each player's choices must be serially independent. That is, intertemporal links between occurrences must be absent. Hence, players' choices must be independent draws from a random process and should not depend on one's own previous play, on the opponent's previous play, or on any other previous actions and outcomes.

Using data on over a thousand penalty kicks during a five year period in three countries, Palacios-Huerta (2003) finds strong support for the two implications of this 2x2 model. We adopt this model and take it to the laboratory. The payoffs we will use in the experiment are:

$$\pi_{LL} = 0.60; \ \pi_{LR} = 0.95; \ \pi_{RL} = 0.90; \ \pi_{RR} = 0.70,$$

which come from a sample of 2,717 penalty kicks collected from professional leagues in Europe during the period 1995-2004.[14] No other field referents are used in the experiment, and no references are made to soccer terminology or any aspect of the natural environment that may trigger any type of psychological motivations.[15] In particular, subjects are not told that the structure of the game corresponds to a penalty kick or that the payoffs correspond to empirically observed probabilities.

The rules of the experiment, which follow as closely as possible O'Neill's (1987), are the following. The players sat opposite each other at a table. Kickers played the

[14]The exact empirical probabilities in the sample are $\pi_{LL} = 0.597$, $\pi_{LR} = 0.947$, $\pi_{RL} = 0.908$, and $\pi_{RR} = 0.698$. The sample includes the 1,417 penalties studied in Palacios-Huerta (2003), which discusses how to treat the few ocassions in which the strategy of "center" may be observed in the soccer field.

[15]As is well known, the choice of parameters can add some field context to experiments. The idea, pioneered by Grether and Plott (1984) and Hong and Plott (1982), is to estimate parameters that are relevant to field applications and take these into the lab.

role of row player and goalkeepers the role of column player. Each held two cards (A and B) with identical backs. A large board across the table prevented them from seeing the backs of each opponent's cards. The experimenter gave them one page with the following instructions (in Spanish), which he then read aloud to them:

> "We are interested in how people play a simple game. You will first play this game for about 15 hands for practice, just to make sure you are clear about the rules and the results. Then, you will play a series of hands for real money at 1 euro per hand. Before we begin, first examine these dice. They will be used at some point during the experiment. They generate a number between 1 and 100 using a 10-face die for the decimal place and another 10-face die for the units. The faces of each die are marked from '0' to '9,' so the resulting number goes from '01' to '99,' where '00' means 100. [The two subjects examine the dice and play with them.] The rules are as follows:
>
> 1. Each player has two cards: A and B.
>
> 2. When I say "ready" each of you will select a card from your hand and place it face down on the table. When I say "turn," turn your card face up and determine the winner. (I will be recording the cards as played).
>
> 3. The winner should announce "I win," and will then receive 1 euro.
>
> 4. Then return the card to your hand, and get it ready for the next round.
>
> I will explain how the winner is determined next. Are there any questions so far?
>
> Now, the winner is determined with the help of the dice as follows:
>
> • If there is a match AA, [row player's name] wins if the dice yield a number between 01 and 60; otherwise [column player's name] wins.
>
> • If there is a match BB, [row player's name] wins if the dice yield a number between 01 and 70; otherwise [column player's name] wins.
>
> • If there is a mismatch AB, [row player's name] wins if the dice yield a number between 01 and 95; otherwise [column player's name] wins.
>
> • If there is a mismatch BA, [row player's name] wins if the dice yield a number between 01 and 90; otherwise [column player's name] wins.
>
> The following diagram may be useful:

$1\backslash 2$	A	B
A	.60	.95
B	.90	.70

Are there any questions?"

Thus, the game was presented with the help of a matrix, and subjects learned the rules by a few rounds of practice. The unique mixed-strategy equilibrium in this game is 0.3636 and 0.4545 for the probability of choosing left for the row and column player respectively. They played 15 rounds for practice and then 150 times for real money, proceeding at their own speed. They were not told the number of hands they will play. If they happened to make an error announcing the winner, the experimenter corrected them.

The typical session lasted about one hour and fifteen minutes, proceeding at about 2 hands per minute. From the perspective of the response times study of Rubinstein (2005) on instinctive and cognitive reasoning, it may be of interest to note that professionals took on average 70.0 minutes, which is 15 percent less time than the average time taken by college students: 81 minutes and 24 seconds. The difference is statistically significant.

2.2.2 Experiment 2: O'Neill (1987)

The design of this experiments closely follows O'Neill's original design. The players sat opposite each other at a table. Each held four cards with identical backs. A large board across the table prevented them from seeing the backs of each opponent's cards. The experimenter gave one page with the following instructions (in Spanish) to the participants, which he then read aloud to them:

> "We are interested in how people play a simple game. You will first play this game for about 15 hands for practice, just to make sure you are clear about the rules and results. Then, you will play a series of hands for money at 1 euro per hand. The rules are as follows:
>
> 1. Each player has four cards: {*Red, Brown, Purple, Green*}.
> 2. When I say "ready" each of you will select a card from your hand and place it face down on the table. When I say "turn," turn your card face up and determine the winner. (I will be recording the cards as played).
> 3. The winner should announce "I win," and will then receive 1 euro.
> 4. Then return the card to your hand, and get it ready for the next round.
>
> Are there any questions?
>
> Now, to determine the winner: [subject 1's name] wins if there is a match of *Greens* (two *Greens* played) or a mismatch of other cards (*Red-Brown* for example); hence, [subject 2's name] wins if there is a match of cards other than *Green* (*Purple-Purple* for example) or a mismatch of a *Green* (one *Green*, one other card)."

Thus, the game was presented *without* the help of a matrix and subjects learned the rules by practice. The payoff structure of the game is:

1\2	*Red*	*Brown*	*Purple*	*Green*
Red	−	+	+	−
Brown	+	−	+	−
Purple	+	+	−	−
Green	−	−	−	+

where the '+' and '−' symbols denote a win by the row and column player respectively. The stage and the repeated games have a unique equilibrium which requires both players to randomize with probabilities 0.2, 0.2, 0.2, 0.4, respectively. Subjects played 15 rounds for practice and then 200 times for real money, proceeding at their own speed. They were not told the number of hands they will play. If they happened to make an error in determining the winner, the experimenter corrected them.

A first difference with respect to O'Neill's design is that the subjects engage in 200 stage games instead of 105. A second difference involves renaming the elements of the action space. Rather than using {*Ace, Two, Three, Joker*}, we use {*Red, Brown, Purple, Green*}, as in Shachat (2002), in order to avoid the previously observed *Ace* bias.[16] Yet, in order to avoid confusion and to facilitate comparison with the literature, actions will be refereed to by the names used in O'Neill's experiment for the remaining exposition of the paper: 1 (*Ace*) for *Red*, 2 (*Two*) for *Brown*, 3 (*Three*) for *Purple*, and *J* (*Joker*) for *Green*. A final difference involves using much greater stage game payoffs (the winner receives 1 euro for a win, that is about 1.30 dollars using the exchange rate at the time the experiment took place, rather than 5 cents), and not giving any initial endowments to the players.

For the experiment with college students, these differences are useful to study the extent to which, relative to the implementation in O'Neill, a larger number of repetitions and much greater number of payoffs may take students in the direction of the unique equilibrium of the game.

The typical session lasted slightly above one hour, proceeding at about 3.3 hands per minute. As in the previous case, professionals took less time than college students (in this case about 11 percent less time on average: 61.2 versus 67.9 minutes). The difference is statistically significant.

[16] See O'Neill (1987) and Brown and Rosenthal (1991).

3 Empirical Evidence

This section is structured as follows. We first describe the evidence from the penalty kick experiment for both the professionals and the college students with no soccer experience, and then the results for O'Neill's experiment for each of these two pools of subjects. In section 4 we discuss the results for the college students with soccer experience in both experiments, and other extensions.

3.1 Penalty Kick Experiment

3.1.1 Professional Soccer Players

Table 1A presents aggregate statistics describing the outcomes of the experiment. In what follows we use the standard notation of L and R instead of A and B. In the top panel each interior cell reports the relative frequency with which the pair of moves corresponding to that cell occurred. In parenthesis are the Minimax relative frequencies and in brackets the standard deviation for the observed relative frequencies under the Minimax hypothesis. At the bottom and to the right are the overall relative frequencies with which players were observed to play a particular card, again accompanied by the corresponding relative frequencies and standard deviations under the Minimax model. Observed and Minimax win frequencies for the row player are reported in the bottom panel.

[Table A1 here]

These aggregate data seem to conform well to the equilibrium predictions. There is a general consistency with the Minimax model in the pattern of observed relative frequencies for each pair of choices, especially for the pair that is played more frequently, RR. As to the marginal frequencies of actions for the players, they are extremely close to the Minimax predictions for the column player. Row players, on the other hand, choose frequencies 0.333 for L and 0.667 for R, which are close to the Minimax predictions but statistically different from them. As to the aggregate row player observed win frequency (0.7947), it is less than one standard deviation away from the theoretically expected value (0.7909).

Data at the individual pair level allow a closer scrutiny of the extent to which Minimax play may be supported for most individual subjects and most pairs of players. Table A2 reports the relative frequencies of choices for each of the twenty pairs in the sample and some initial tests of the model.

[Table A2 here]

The Minimax hypothesis implies that the choices of actions represent independent drawings from a binomial distribution where the probabilities of L are 0.363 and 0.454 for the raw and column players, respectively. We should then expect a binomial test of conformity with Minimax play to reject the null hypothesis for 2 players at the 5 percent significance level, and 4 players at the 10 percent level. The results show that indeed these are precisely the number of rejections at those confidence levels.

Clearly, these initial findings may be taken as consistent with the hypothesis that professional soccer players play according to the equilibrium of the game. Yet, they lend only partial support at this point. One reason is that equilibrium behavior also implies that action combinations should be realizations of *independent* drawings of a multinomial distribution. In principle, it might well be the case that the marginal frequencies conform well to the equilibrium strategies while at the same time the players's actions are highly correlated.

In order to test whether the players' actions are correlated we perform two tests. First, Minimax play implies that action combinations are realizations of independent drawings from a multinomial distribution with probabilities 0.165 for LL, 0.198 for LR, 0.289 for RL and 0.347 for RR. Table A2 reports the relative frequencies of each combination of actions for each of the twenty pairs in the sample. Using the corresponding absolute frequencies along with their Minimax probabilities, we can then test the joint hypothesis that players choose actions with the equilibrium frequency *and* that their choices are stochastically independent. A Chi-square test for conformity with Minimax play based on Pearson's goodness of fit with 3 degrees of freedom produces the p-values reported in the last column of the table. Under the Minimax play we would expect to reject the null hypothesis for 1 and 2 pairs at the 5 and 10 percent significance levels. We find 0 and 2 rejections, respectively, virtually what we would expect.

The second test is the one Brown and Rosenthal (1990) devised to check for *contemporaneous correlation* in players' choices. The results are shown in Table A3.

[Table A3 here]

The first column reports the observed win percentage. The second reports the expected win percentage, using observed card frequencies, under the assumption that the players' choices are i.i.d. drawings from a pair-specific stationary binomial distribution. Column three shows the effect of chosen mixtures, which is computed as the difference between the observed row player winning percentage and the Minimax value of 0.7909. The difference between the first two columns measures the contribution of correlated play to observed row-player winning percentages, and it is reported in the fourth column. It is apparent that the average absolute contributions of both the mixture and the correlated play effects are extremely small. The Chi-square statistics for the significance of each of these effects are minuscule, the probability values for which are virtually unity. These results may be taken as consistent with the hypothesis that professional players play according to the equilibrium of the game.

We take from this initial evidence that, even though the observed aggregate frequency for the row players is slightly different from the equilibrium predictions, the hypothesis of Minimax play gets substantial support in the tests of the binomial and multinomial models for observed choice frequencies, and in the tests of independence. Next, we turn to testing more closely the implications of the equilibrium of the game.

i. Winning Rates and the Distribution of Play

Minimax play implies that the success probabilities of each action should be the same for each player, and equal to 0.7909 for the row player and 0.2090 for the column player. Further, when combined with the equilibrium strategies, we can obtain the relative success-fail frequencies associated with each action in equilibrium.

Table A4 reports the relative frequencies of action-outcomes combinations observed for each of the row and column players in the sample. Using the absolute frequencies corresponding to these entries, we can then implement a Chi-square test of conformity with Minimax play. This test would be identical to the one performed in Table A2 if it were not for the fact that the success rate is determined not only by the choice of strategies but also by the realization of the dice.

[Table A4 here]

The results of the test show that the null hypothesis is rejected for no player at the 5 percent significance level, and for 3 players at the 10 percent significance level, in both cases fewer than the expected number of rejections, 2 and 4 respectively. Hence, at the individual level the hypothesis that scoring probabilities are identical across strategies and to the equilibrium rate cannot be rejected for most players at conventional significance levels.

With regard to whether behavior at the aggregate level may be considered to be generated from equilibrium play, this idea may be evaluated by testing the joint hypothesis that each one of the experiments is simultaneously generated by equilibrium play. The test statistic for the Pearson joint test is simply the sum of the individual test statistics for each type of players. Under the null hypothesis, it is distributed as a χ^2 with 60 degrees of freedom for the set of row players and likewise for the set of column players. We find that the Pearson statistics are 40.002 and 32.486, with associated p-value above ninety percent in both cases.[17] Hence, the null hypothesis that the data for all players were generated by equilibrium play cannot be rejected at conventional significance levels.

We interpret these individual and aggregate results as consistent with the hypothesis that these subjects equate their strategies' payoffs to the equilibrium success rates.

[17] The test statistics for the raw and column players may not be added given that within each pair the players' success rates are not independent.

II. The Serial Independence Hypothesis

The second testable implication of equilibrium play is that a player should randomize using the same distribution at each stage of the game. This implies that players' choices are serially independent. The work on randomization is extensive in the experimental economics and psychological literatures. Interestingly, this hypothesis has *never* found support in a laboratory setting. In particular, when subjects are asked to generate random sequences their sequences often have negative autocorrelation, that is, individuals exhibit a bias against repeating the same choice (see Bar-Hillel and Wagenaar (1991), Rapoport and Budescu (1992), Rapoport and Boebel (1992) and Mookherjee and Sopher (1994)).[18] This phenomenon is sometimes referred to as the "Law of Small Numbers" (Tversky and Kanheman (1971) and Camerer (1995)). The only possible exception that we are aware of is Neuringer (1986) who explicitly taught some subjects to choose randomly after several hours of training by providing them with detailed feedback from previous blocks of responses in the experiment. These training data are interesting in that they suggest that experienced subjects might be able to learn to generate randomness. In our case, however, subjects have accumulated their experience in an entirely different environment: a soccer field. Moreover, professional soccer players rarely take penalty kicks in the field in rapid succession, as they are asked to in the experiment. Instead, there is often a substantial time delay, typically weeks, between subsequent penalties.[19] Whether their experience on randomization in the field in circumstances where repetitions are not taken in rapid succession is useful to generate random sequences in a laboratory setting where stage games are repeated in rapid succession is the question to which we turn next.

We consider the following tests of serial independence previously performed in the literature:

A. Runs Tests. Consider the sequence of strategies chosen by player i in the order in which they occurred $s^i = \left\{s_1^i, s_2^i, ..., s_{n^i}^i\right\}$, where $s_x^i \in \{L, R\}$, $x \in [1, n^i]$, $n^i = n_L^i + n_R^i$, and n_R^i and n_L^i are the number of R and L choices made by player i. Let r^i denote the number of runs in the sequence s^i. A run is defined as a succession of one or more identical symbols which are followed and preceded by a different symbol

[18] Slonim, Erev and Roth (2003) find evidence of positive autocorrelation in various zero-sum 2x2 games.

[19] Most players are not involved in more than 15-20 penalty kick situations per season, that is in about one such situation every two weeks. It could be argued that these time lags between penalties might inhibit the memory of past realizations, which would in turn help players randomize correctly. However, this conjecture may not be correct. The reason is that every penalty kick that is taken in professional leagues is televised and a common practice among professional players is to keep written records on their opponents' behavior (Anthony, 2000). Keeping records may then induce, perhaps, even better memory than if penalties were shot in rapid succession. Penalty shoot-outs that take place in elimination tournaments to break ties are an exception in which kicks occur in a short span of time. Even in these cases, however, only five penalties are typically taken and these involve five *different* kickers facing the same goalkeeper.

or no symbol at all. Let $f(r^i; s^i)$ denote the probability that there are exactly r^i runs in the sequence s^i. More precisely, $f(r; s^i)$ is the probability of r runs of $n^i = n^i_L + n^i_R$ action choices, n^i_L left and n^i_R right, when choices are independently and equally distributed. Gibbons and Chakraborti (1992) show that this probability is given by

$$
\begin{aligned}
f(r;s^i) &= \begin{cases} 2\binom{n^i_L-1}{r/2-1}\binom{n^i_R-1}{r/2-1}/\binom{n^i_L+n^i_R}{n^i_L} & \text{if } r \text{ is even} \\ \left(\binom{n^i_L-1}{(r-1)/2}\binom{n^i_R-1}{(r-3)/2}+\binom{n^i_L-1}{(r-3)/2}\binom{n^i_R-1}{(r-1)/2}\right)/\binom{n^i_L+n^i_R}{n^i_L} & \text{if } r \text{ is odd} \end{cases} \\
\text{for } r &= 2,3,\ldots,n^i_L+n^i_R.
\end{aligned}
$$

Letting $F(r; s^i) = \sum_{k=1}^{r} f(k; s^i)$ denote the probability of obtaining r or fewer runs the null hypothesis will then be rejected at the 5 percent confidence level if the probability of r or fewer runs is less than 0.025 or if the probability of r or more runs is less than 0.025; that is, if $F(r; s^i) < 0.025$ or if $1 - F(r-1; s^i) < 0.025$.[20] The results of these tests are shown in Table A5.

[Table A5 here]

We find that the null hypothesis of serial independence is rejected for very few players at conventional significance levels: 2 players at the 5 percent significance level and 4 players at the 10 percent level, precisely the expected number of rejections in both cases under the null hypothesis. These results indicate that the hypothesis that professional soccer players generate random sequences cannot be rejected according to this test. They neither switch strategies too often nor too little, and the number of rejections is remarkably consistent with the theory. This behavior, therefore, is in sharp contrast with the overwhelming experimental evidence from the psychological and experimental literatures mentioned earlier.[21] In this case it shows that years of experience in the field is quite valuable, even if it comes from situations where repetitions are not taken in rapid succession, and from circumstances that are vastly different from those they find in the laboratory.

B. Logit Equation for Individual Players. Brown and Rosenthal (1990) suggest using a logit model to study whether past choices and outcomes play a role

[20] The critical values for the rejection of the hypothesis can also be found from the normal approximation to the null distribution that the authors offer.

[21] Note also that the values in columns $F(r^i; s^i)$ and $F(r^i - 1; s^i)$ seem to be uniformly distributed in the $[0, 1]$ interval. Formally, the joint hypothesis that each of the forty experiments is serially independent may be tested following the suggestion in Walker and Wooders (2001). Given that a Kolmogorov-Smirnov (KS) test cannot be applied directly to $F(r^i; s^i)$ and $F(r^i - 1; s^i)$ because these values are neither identically nor continuously distributed, a KS goodness-of-fit test can be developed by constructing a random draw d^i from the uniform distribution $U[F(r^i - 1; s^i), F(r^i; s^i)]$ for each player i. Under the null hypothesis of serial independence d^i is distributed as a $U[0, 1]$. After performing ten thousand trials with such random draws for each player, the average p-value of the KS test statistic that compares the cumulative distribution of the realized values d^i with the uniform distribution is 0.507 with a standard deviation of 0.166. Hence, the hypothesis that each of the forty experiments is serially independent cannot be rejected.

in determining current choices. We follow the formulation suggested in Slonim, Erev and Roth (2003). The dependent variable is a dichotomous indicator of the choice R. The independent variables are a lagged indicator of the same choice, an interaction between that indicator and whether the subject won in the past round, an interaction between the lagged alternative choice and whether the subject won in the past round, a lagged indicator of the opponent's same choice, and an indicator of the opponent's contemporaneous choice R^*. The results are shown in Table A6.

[Table A6 here]

Consistent with the evidence from the runs tests, the main finding is that the null hypothesis that *all* the explanatory variables are jointly statistically insignificant (hypothesis #1), i.e., that professional subjects follow a stationary binomial choice process, can be rejected for only 2 players at the 5 percent level and 4 players at the 10 percent level.

The table also reports the tests of other hypotheses of interest. Hypothesis #2, which studies whether one's past choices significantly help to determine current choices, is rejected for only 2 and 5 players at the 5 and 10 percent levels respectively. Not surprisingly, it is rejected for some of the players that in the previous table had either a high or a low number of runs.[22] The "reinforcement" hypotheses #3 and #4, which evaluate whether subjects are more likely to repeat an action in round t if they won in round $t-1$ ($\beta_3 > 0$, $\beta_4 < 0$) or less likely ($\beta_3 < 0$, $\beta_4 > 0$), only find support in at most two cases at the 5 percent level. The results for hypothesis #5 support the idea that players believe that his opponent is using a stationary choice rule in every case except 3 and 4 at the 5 and 10 percent levels. Lastly, the tests of hypothesis #6 show no effect of contemporaneous opponent's choice on one's current choice for any player except one.

These results are consistent with the previous test of serial independence and indicate that the choices of most players are unrelated to their own previous choices, to opponents' previous choices, and to past outcomes. We thus take the results of the two tests of randomness as consistent with the hypothesis that the strategies followed by professional soccer players are serially independent. As such, this evidence represents the first time that individuals have been found to display statistically significant serial independence in a strategic game in a laboratory setting. Jointly with the evidence supporting the hypothesis that subjects equate payoffs across strategies and to the equilibrium success rates, these results also represent the first time that any subjects reach a predicted equilibrium in the laboratory in games where players are predicted to choose probabilistic mixtures. Hence, laboratory findings are entirely reliable for predicting field behavior for these subjects.

[22] For row player #5 and column players #3 and #5 the reason is $\beta_2 < 0$, and for row player #7 and column player #11 it is $\beta_2 > 0$.

3.1.2 College Students

The results for this subject pool are presented in a way that parallels the presentation of the evidence for the professional soccer players. Table B1 presents aggregate statistics describing the aggregate outcomes of the experiment.

[Table B1 here]

Interestingly, the aggregate data for these players also seem to conform to the equilibrium predictions quite well. There is a broad consistency of the observed relative frequencies with those implied by the Minimax model, especially for the diagonal pairs of choices. Moreover, as in the case of professionals, the observed aggregate win frequency for the row player (0.7877) is also below one standard deviation away from the expected value. Despite these appearances, however, a closer look quickly reveals that observed behavior is far from the Minimax predictions. For instance, observed marginal frequencies for both the row and column players are substantially different from the predicted values, somewhere between 4 to 6 standard deviations away from them.[23] An interesting aspect worth noting is that *both* players choose very similar frequencies, roughly 0.40 for L and 0.60 for R. This suggests the possibility that these subjects, contrary to the way professionals appear to perceive the game, may not appreciate the slight differences in payoffs in the off-diagonal elements of the payoff matrix, differences that induce players to adopt strategies different from the opponent.

The rejections of Minimax play are even more apparent in Table B2, which reports the marginal frequencies for each player and the relative frequencies of choices at the pair level.

[Table B2 here]

First, the binomial test for conformity with Minimax play indicates that the model is rejected for 6 and 22 players at the five and ten percent levels respectively. This excessively high amount of rejections, three and more than five times greater than those predicted by the equilibrium of the game at those levels, may be taken as an indication that we are going to find substantial deviations from equilibrium play in the subsequent tests of the Minimax hypothesis. Indeed, using the absolute frequencies corresponding to the observed joint choices reported in the table and their associated Minimax probabilities, a Chi-square test for conformity with Minimax play indicates that the model is rejected for 6 and 9 pairs at the five and ten percent levels of significance when we would only expect 1 and 2 rejections, respectively, under the null hypothesis.

[23] The Chi-square test for the conformity with Minimax play based on Pearson goodness of fit has a p-value of 2×10^{-13}, which is minuscule compared to professionals (4.8 percent) and decisively rejects the Minimax model at conventional significance levels.

The case of pairs #12 and #20 is interesting. Although the marginal frequencies with which the players choose each action are not statistically different from the equilibrium strategies, their joint behavior rejects the equilibrium multinomial model. As can be seen from the data, their joint behavior is highly correlated in that they tend to choose main diagonal entries too frequently. As we did for professionals, this aspect is studied further by decomposing the players' observed behavior into a mixture effect and a correlation effect. The results are in Table B3.

[Table B3 here]

We find that the average absolute contribution of the mixture effect is extremely small: the Chi-square statistic is 0.1138, with a probability value that is virtually one. As to the average absolute contribution of the correlated play effect, the mean absolute value is 0.0279, with a Chi-square statistic of 22.3481 and a probability value of 0.3219. Under conventional standards, this p-value is clearly high, so we may not conclude that correlated play is the dominant determinant of the difference between observed row-player winning percentages and the Minimax predicted value. However, it does suggest the possibility that for some college students this might be a relevant effect. At least, relative to professional soccer players (Chi-square: 3.2344, p-value: 0.9999), this p-value allow us to suspect that several subjects may deviate from independence play.

Overall, we take the excessively high amount of rejections we find in these tests as indicating substantial deviations from equilibrium play. Next we turn to testing the Minimax predictions more closely.

i. Winning Rates and the Distribution of Play

In Table B4 we test whether the distribution of play we observe is equal to the equilibrium distribution using the success rates of each action for each player.

[Table B4 here]

Using the absolute frequencies corresponding to each action-outcome combination, a Chi-square test shows that the Minimax multinomial model is rejected for 9 players at the five percent significance level and 13 players at the ten percent level. These rejections vastly exceed the expected number of rejections under the hypothesis of Minimax play, 2 and 4 respectively. Thus, at the individual level the hypothesis that scoring probabilities are identical across strategies and equal to the equilibrium strategies can be rejected for an excessively high number of players at conventional significance levels.

With regard to aggregate behavior, the sum of the individual test statistics of each type of player under the null hypothesis is distributed as a χ^2 with 60 degrees of freedom. For the row players the joint test statistic is 108.652 and for the column players 113.102, with associated p-values close to zero in both cases. Hence, the null hypothesis that the data for all players were generated by equilibrium play is strongly rejected at conventional and non-conventional significance levels.

These results, therefore, are far from equilibrium behavior and highly different from those obtained with professional soccer players.

II. The Serial Independence Hypothesis

The second testable implication is that a player should randomize by means of the same probability distribution at each stage of the game. We implement the same two tests of serial independence implemented for professional players.

A. First, the results of the runs tests of serial independence shown in Table B5 confirm earlier suspicions.

[Table B5 here]

The null hypothesis of serial independence is rejected for 7 players at the 5 percent significance level, more than three times the number of expected rejections, and for 13 players at the 10 percent when we would only expect 4 rejections. These findings indicate that college subjects are not able to generate random sequences. Hence, they are consistent with an extensive experimental evidence in the literature and drastically different from the behavior of professional soccer players observed earlier.[24] Also consistent with past evidence is the fact that in most cases the reason is an excessive number of alternations.

B. The results of the logit equation for each player for the choice of R are shown in Table B6.

[Table B6 here]

The main finding is that the null hypothesis that college students follow a stationary binomial choice process in this experiment is rejected for an excessive number of subjects: 12 players at the 5 percent level and 14 players at the 10 percent level. Other hypotheses of interest are also frequently rejected. For instance, hypothesis #2, which tests whether or not one's past choices contribute to determining current choices, is rejected for 6 players at the 5 percent level and 8 players at the 10 percent level. This suggest that a failure to play independently of one's past choices is an

[24] As in the case of professionals, we also implemented a KS goodness-of-fit test by constructing a random draw d^i from the uniform distribution $U\left[F\left(r^i-1;s^i\right),F\left(r^i;s^i\right)\right]$ for each player i, which under the null hypothesis of serial independence is distributed as a $U[0,1]$. After performing ten thousand trials with such random draws for each player, the average p-value of the KS test statistic is 0.0002 with a standard deviation of 0.0116. Hence, the hypothesis that each of the forty experiments is serially independent can be rejected.

important reason behind many subjects' inability to generate a stationary choice process. Consistent with the runs tests, in most of these cases the reason is a significant preference for alternation ($\beta_2 < 0$). With regard to the reinforcement hypothesis #3 and #4, we find that reinforcement contributes to the non-stationary choice process in some but not a very large number of cases as well, while the results for hypothesis #5 support the idea that 8 players at the 10 percent level do not consider that his opponent is using a stationary choice rule. The evidence from these four hypothesis, therefore, shows that for almost half of the sample (19 players) at least one of these hypothesis is rejected at the 10 percent level. Lastly, hypothesis #6 shows a significant effect of contemporaneous opponent's choice on one's current choice in precisely five pairs of players at the 10 percent level.

These results are consistent with the runs tests of serial independence: the choices of many players are related to their own previous choices, outcomes, and those of the opponent in various ways, which contribute to generating non-stationary choice processes. Consequently, the results of the tests of serial independence decisively indicate that individuals display statistically significant serial dependence. Together with the results in the tests of equality of winning probabilities, we can then conclude that the Minimax model cannot be supported for college students.

3.2 O'Neill Experiment

The differences between professional soccer players and college students are substantial in the penalty kick experiment. Professionals behave consistent with the equilibrium of the game while college students far from it. In this section we examine a different zero-sum game in an attempt to study whether the experience that professional players have accumulated in the field is useful in laboratory situations that do not resemble any previously encountered situation. We implement the same tests as in the penalty kick experiment.

3.2.1 Professional Players

Table C1 presents aggregate statistics describing observed relative frequencies for each pair of moves and each card. Minimax relative frequencies are in parenthesis, and their standard deviations under the Minimax hypothesis are in brackets. The bottom panel reports the observed win frequencies for the row player.

[Table C1 here]

These aggregate data seem to conform remarkably well to the equilibrium predictions. In fact, there is a striking consistency of the observed relative frequencies with those implied by the Minimax model. Relative frequencies for pairs of plays involving non-jokers are in the neighborhood of 0.04, while relative frequencies for pairs involving one joker and for the pair involving the two jokers are in the neighborhood of 0.08 and 0.16 respectively. The aggregate row player win frequency (0.3945) is less than one standard deviation away from the expected value (0.40). Also, a Chi-square test for the conformity with Minimax play based on Pearson goodness of fit indicates that the Minimax model cannot be rejected at conventional significance levels. It yields a statistic of 7.873 whose p-value is above ninety percent. As to the marginal frequencies of actions for the row and column players, they are extremely close to the Minimax predictions. In every case, they are less than one standard deviation away.

This evidence, however, while highly suggestive, does not mean that Minimax play can conclusively be supported by the data. Indeed, Brown and Rosenthal (1990) already found a substantial degree of conformity in O'Neill's (1987) experiment in the aggregate data only to find in subsequent tests that the Minimax hypothesis could not possibly be supported.

Table C2 reports the marginal choice frequencies observed in the data for each player, and the results of the tests of the Minimax model at the individual player, pair, and card-player levels for these frequencies.

[Table C2 here]

Minimax play indicates that the multinomial model for all cards chosen by row players, column players, and for both players should be rejected for 1 pair at the 5 percent level and for 2 pairs at the 10 percent level in each of these three models. The Minimax binomial model for a given card indicates that we should expect 8 and 16 rejections at these significance levels respectively. As in the penalty experiment, it is remarkable that in virtually every case these are the precise number of rejections that are found. Furthermore, this evidence is strikingly dissimilar from Brown and Rosenthal's (1990) analysis of O'Neill's data with college students.

As indicated in the penalty experiment, however, the fact that the marginal frequencies with which players choose each card seem to correspond to the equilibrium strategies lends only partial support to the model. The reason is that in principle the same marginal frequencies may be obtained as a result of a correlated strategy. To check for contemporaneous correlation we perform Brown and Rosenthal's decomposition of row players' winning rates into a mixture and a correlated effect.

[Table C3 here]

The results in Table C3 show that the average absolute contribution of both the mixture and the correlated play effects are extremely small. The Chi-square statistics for the joint significance of each of these effects are minuscule, with probability values close to one. These results suggest that professional soccer players do not show contemporaneous correlation in their choices.

We turn next to a closer examination of the implications of the equilibrium of the game.

i. Winning Rates and the Distribution of Play

Table C4 tests the null hypothesis that the success probabilities for both players are identical across strategies and equal to the equilibrium probabilities. As in Walker and Wooders (2001) analysis of O'Neill's data, we aggregate actions *1, 2,* and *3* into a single *non-Joker* action. We then implement the corresponding χ^2 test of conformity with Minimax play. The tests have three degrees of freedom given that the game being played is known. In other words, as the success probabilities and choice frequencies in equilibrium are known, there is no need to estimate any parameter.[25] The table also indicates the rejections that are obtained when the test is implemented for the individual choices of cards, that is when *1, 2, 3*, and *J* are treated on an individual basis.

[Table C4 here]

The results show that for the choice of *Joker* and *non-Joker* the null hypothesis is rejected for 3 players at the 5 percent significance level and for 6 players at the 10 percent significance level. The number of rejections when the test is implemented for the individual cards is 3 and 4 at these levels respectively, which are almost precisely the number of rejections to be expected according to the null hypothesis.

With respect to whether behavior at the aggregate level can be considered to be generated from equilibrium play, the test statistic for the Pearson joint test for all row players is 53.351 with an associated p-value of 0.715, and for all column players 55.122 with an associated p-value of 0.654. Hence, the null hypothesis that the data for all players were generated by equilibrium play cannot be rejected at conventional significance levels.

ii. The Serial Independence Hypothesis

Another testable implication is that players' choices are serially independent. We implement two tests:

A. Runs Tests. As in Walker and Wooders (2001) this test is implemented for the choice of *Joker* and *non-Joker* cards. Table C5 shows that the null hypothesis of serial independence is rejected for 2 and 4 players at the 5 and 10 percent significance levels. According to the theory, this is precisely the number of rejections that we should expect at these levels.

[Table C5 here]

[25] In field data as in soccer penalty kicks or in tennis serves, and contrary to laboratory experiments, the underlying game is not known. Hence, the success probabilities for each player must be estimated by maximum likelihood. Walker and Wooders (2001) test for the equality of two distributions in O'Neills experimental data also using the maximum likelihood estimate for the success probabilities for each player, rather than the equilibrium rate. Using the equilibrium rates in their data, however, yields similar results to the ones they report.

These findings, therefore, also support the hypothesis that professional soccer players are able to generate random sequences in the laboratory.[26]

B. Logit equation for each player. To study whether or not past choices have a role in determining current choices, we estimate the logit equation for each player suggested by Brown and Rosenthal (1990). The dependent variable is a dichotomous indicator of the choice J. The independent variables are first and second lagged indicators for both players' past choices, first and second lags for the product of their choices, and an indicator for the opponent's current choices. The results are shown in Table C6.

[Table C6 here]

The main finding is that the null hypothesis that *all* the explanatory variables are jointly statistically insignificant (hypothesis #1) can be rejected for relatively few players at both the 5 and 10 percent levels. The results for hypotheses #2 to #5 indicate that they are rejected for no more than 3 players at the 5 percent level, typically row players #2 and #11 and column player #2. At the 10 percent level, hypothesis #2 is the only one that shows a slightly greater number of rejections, while the reinforcement of successful actions (hypotheses #3 and #4) is significant at this level for very few players. Lastly, there is no evidence of correlation with opponent's past actions (hypothesis #5) or with opponent's contemporaneous actions (hypothesis #6) for any player except for three row players. These results indicate that the choices of most players are unrelated to their own previous choices and outcomes, and to opponents' previous choices and outcomes.

We take the results of these three tests of randomness as consistent with the hypothesis that the strategies followed by professional soccer players are serially independent. As such, this class of subjects continues to display statistically significant serial independence in their choices within a laboratory setting. As indicated in the penalty kick experiment, this behavior is sharply different from the overwhelming experimental evidence reported in the psychological and experimental literature on randomization, which consistently finds that subjects, typically college students, generate sequences that exhibit negative autocorrelation. The interesting additional aspect is that, in this case, professionals are involved in a zero-sum game that is entirely different in terms of number of strategies and payoff structure from the penalty kick experiment or other situation they may have found in the field.

[26]The KS goodness-of-fit test constructed by performing ten thousand random draws d^i from the uniform distribution $U\left[F\left(r^i-1;s^i\right),F\left(r^i;s^i\right)\right]$ for each player i, yields an average value of the p-value of the test statistic that compares the cumulative distribution of the realized values d^i with the uniform distribution of 0.420 with a standard deviation of 0.117. Hence, the hypothesis that each of the forty experiments is serially independent cannot be rejected.

3.2.2 College Students

In principle, it is conceivable that it is the greater stage payoffs that we offer and the greater number of repetitions that we undertake in the experiment relative to previous implementations of the experiment in the literature, and *not* the field experience of the subjects, that are causing the consistency with the Minimax hypothesis. Thus, we turn next to the study of college students under identical circumstances to those faced by professionals.

The results are presented in Tables D1 to D6 in a way that parallels the presentation of the empirical evidence for the professional soccer players. They can be summarized as follows.

We find that the general results in Brown and Rosenthal (1991), Walker and Wooders (2001) and Shachat (2002) with O'Neill's experiment are replicated here. Even though aggregate frequency data does not seem too far from equilibrium behavior, the Minimax hypothesis is decisively rejected in virtually every test we implement. Observed aggregate row player win percentage is more than one standard deviation away from the predicted value (Table D1). Observed card, player, and card-player choices reject the different Minimax multinomial models for all cards, as well as the Minimax binomial model for a given card, in an excessively large number of cases (Table D2), and correlated play effect cannot be argued to be the dominant effect at conventional significance levels (Table D3). Individual Pearson tests for the equality of winning rates to the equilibrium one are also rejected for a very high number of subjects; at the aggregate level the joint hypothesis is decisively rejected for all row players and all column players as well, both when cards are treated as NJ and J, and when they are treated on an individual basis (Table D4). There is strong evidence that too many players relative to the Minimax predictions exhibit statistically significant serial dependence in the runs tests (Table D5), in fact about three times the number of rejections observed for professional players. Finally, the logit equation for individual players reveals that for 13 players in the sample the hypothesis of stationary binomial choice process can be rejected at the 10 percent level (Table D6).

As in the penalty kick experiment, these findings are in sharp contrast with those obtained for professional soccer players. These results also testify to the robustness of previous findings in the literature. Although we use much greater monetary incentives and more repetitions than in O'Neill's original experiment, and we do find improvements in the behavior of college students from the perspective of equilibrium (see Table F in the next section for a comparison), the Minimax model continues to be rejected decisively. Given that the circumstances of the experiment are identical for college students and professional players, and that professionals behave consistent with the equilibrium of the game, the results indicate that field experience is important and does transfer to this zero-sum game as well.

4 Additional Evidence

The findings that professional soccer players play according to equilibrium both in the penalty kick and in O'Neill experiments, while college students are far from equilibrium behavior in both experiments, contributes to dissecting the possibly fundamental characteristics of laboratory experiments as a source of predictions for the real world in the class of strategic games we study. None of the potential drawbacks and limitations associated with the controlled and artificial environment that represents a laboratory seem to induce professional players to play any differently from the way they play in the field (in the penalty kick experiment) and from the equilibrium (in both games).

Additional evidence has been obtained to study various aspects further:

1. We have first studied the robustness of some of the results with smaller samples of subjects. For instance, in the penalty kick experiment we have used payoffs that are entirely different from the scoring probabilities occurring in the field. We have also studied the behavior of professional players where kickers in the soccer field play the role of goalkeepers in the laboratory and vice versa. Although care should be exercised here since our sample sizes are smaller, we find that none of these modifications of the experimental procedures seem to cause any significant changes in the basic results obtained earlier: professional soccer players continue playing consistent with the equilibrium predictions while college students do not.[27] In some sense these results may not be surprising. Professional players are involved in several situations in the soccer field where they must randomize. Yet, few of them are the "designated penalty-kick taker" for their team, and our sample includes few of those. Sure enough, our kickers and goalkeepers have been involved in hundreds of penalty kicks throughout their lives. But it seems likely that it is their general experience randomizing rather than their specific experience in penalty kick situations what helps them reach the equilibrium in the laboratory. This is also consistent with the evidence we obtained for professionals in O'Neill's experiment as they have never been exposed to that precise situation.

2. We have also implemented two other tests of time independence:

(i) First, we have considered the test suggested in Shachat (2002), where the independent variable is the joint realization of a *pair* of strategies using one and two lags. Not surprisingly, we find that there is still substantial serial correlation among standard college students but not among professional and amateur soccer players.

(ii) Second, we have pooled all 40 subjects for each experiment and class of players and estimated a binary choice dynamic panel data model with predetermined endogenous variables and unobserved individual heterogeneity. As is known, in fixed-effect models parameter estimates can be biased and inconsistent when the explanatory variables are predetermined as opposed to strictly exogenous (see Arellano and

[27] These results are available upon request.

Honoré (2002) for a review).[28] Recently, however, Arellano and Carrasco (2003) have developed a *consistent* random effects estimator where: (a) explanatory variables are predetermined but not strictly exogenous, and where (b) individual effects may be correlated with explanatory variables. Thus, in order to control for the potential effect of state dependence caused by past choices and outcomes appropriately in our setting, we estimated this semi-parametric, dynamic random-effects panel data model. The results confirm previous findings. No lagged endogenous variables (past own and opponent's choices and outcomes alone or interacted) are significant for professional and amateur players, while negative autocorrelation and positive reinforcement significantly characterize the behavior of college students.

3. The important differences among subject pools open up various avenues for further research. For instance, it may be of interest to study the extent to which field experience at the *professional* level is necessary to reach the predicted equilibrium. As indicated earlier, we have pursued this question by recruiting subjects drawn from the same pool of college students as the students recruited previously, except that they were required to be currently playing in one of the official amateur senior regional leagues, including Tercera Division, described in Section 2.[29] Playing in these leagues is still quite competitive. Amateur teams practice as often as professionals and have exactly the same 9-month playing schedule. Players in these leagues began playing soccer as early as those that became professional. Hence, conditional on age, they have roughly the same years of field experience. They simply are not as skilled as professional players in the many different aspects of the game.

We implement both the penalty kick experiment and O'Neill's experiment for these subjects. In order to conserve space, rather than showing the corresponding tables, we make them available in an appendix (Palacios-Huerta and Volij, 2006), and just report the main results of each of the tests. Table E shows the results for the penalty kick experiment and Table F for O'Neill's experiment. To facilitate the comparison of the results, we also include the results obtained with professional players and college students with no soccer experience presented earlier. In Table F, in addition, we include the results of Pearson's tests of equality of winning rates when, rather than using the equilibrium value, we use its maximum likelihood estimate,[30] and the original results of O'Neill's experiment reported in Brown and Rosenthal (1990) and Walker and Wooders (2001).

[28] While in linear models with additive effects the standard response is to consider instrumental-variables estimates that exploit the lack of correlation between lagged values of the variables and future errors in first differences, in non-linear models very few results are available. A difficulty is that the regularity conditions for conditional maximum likelihood estimation of a fixed effects logit model are not satisfied in the presence of a lagged dependent variable when the error term at t is not independent of explanatory variables at $t-1$.

[29] There are no statistical differences in the distributions of demographic characteristics such as age and years of education between these two different pools of college students.

[30] This is the procedure that Walker and Wooders (2001) follow in their reanalysis of O'Neill's data. It implies that the tests at the player level have 1 degree of freedom instead of 3.

[Tables E and F here]

We find that the behavior of these subjects adheres in many cases almost as closely as the behavior of professionals to the equilibrium predictions, and sometimes even slightly better. As such they differ greatly from the way the standard pool of college students behave.

These results indicate that years of field experience playing soccer, a game that offers several opportunities to behave strategically in zero-sum situations, are a critical determinant of behavior in the laboratory.

5 Concluding Remarks

This paper has taken advantage of three distinct features: (i) there is a precisely defined strategic situation played in the soccer field whose formal structure can be reproduced in the laboratory, (ii) this situation involves mixed-strategy interaction between subjects and has a unique individually rational payoff vector, (iii) professional subjects play in a real life setting according to the equilibrium of this game. These characteristics are helpful to design a first artefactual field experiment in mixed-strategy interactions that helps us to isolate the role of "laboratory context" and that allows us to compare field and laboratory behavior.

We find that field experience transfers from the familiar soccer field to the highly unfamiliar laboratory when subjects play a game that is formally identical to a game situation that they find in natural circumstances. Field experience is also valuable to reach the equilibrium in a zero-sum game previously studied in the literature that subjects have never faced before.

These results may have theoretical, methodological and cognitive implications:

In terms of the theory, as subjects reach a predicted equilibrium in the laboratory, the theoretical concept of *equilibrium* of the game may have greater predictive power than previously considered, even in artificial settings such as a laboratory.

From a methodological perspective the results are relevant to the extent that the data that are typically used to inform game theory, and increasingly theoretical developments in other areas in economics and social sciences as well, often comes from laboratory environments. In this sense, the insights obtained in the laboratory with the pool of subjects that we would be interested in studying empirically in the field are perfectly applicable for predicting field behavior.

Lastly, from a cognitive perspective our findings are consistent with the idea that skills have been learned unconsciously and are active in the solution of the games we have studied. In this sense, "the demonstrated capacity of motivated subjects to find equilibrium outcomes by repeated interaction in market experiments without cognitive awareness of this capacity" emphasized in Smith (2005) and other authors is supported, for the first time to our knowledge, in situations requiring use of mixed strategies. From this perspective, Camerer, Loewenstein and Prelec (2005) discuss neurological evidence showing how as subjects gain experience with certain games (Haier et al, 1992) and situations (Lo and Repin, 2002) "the brain becomes more streamlined, concentrating in regions that are specialized in processing relevant to the task ... gradually shifting processing toward regions and specialized systems that can solve problems automatically and efficiently." We cannot disregard the idea that years of field experience in different zero-sum strategic situations, not only in penalty kicks, have had these effects in professional soccer players. An alternative we cannot discard either is that those players that became professional were born with greater aptitude for playing strategic zero-sum games than other subjects.

REFERENCES

Anthony, A. (2000), *On Penalties.* Yellow Jersey Press, Random House, London.

Arellano, M. and R. Carrasco (2003), "Binary Choice Panel Data Models with Predetermined Variables," *Journal of Econometrics* 115, 125-157.

Arellano, M. and B. E. Honoré (2002), "Panel Data Models: Some Recent Developments." In *Handbook of Econometrics,* vol. 5, chapter 53. Edited by J. E. Heckman and E. Leamer, North Holland.

Aumann, R. J. (1987), "Game Theory," in *The New Palgrave: A Dictionary of Economics*, volume 2, edited by J. Eatwell, M. Milgate, and P. Newman, pp. 460-482. MacMillan, London.

Azar, O. H. and M. Bar-Eli (2005), "Do Soccer Players Play the Mixed-Strategy Nash Equilibrium?" Mimeo, Ben-Gurion University.

Bar-Hillel, M. and Wagenaar, W. A. (1991), "The Perception of Randomness," *Advances in Applied Mathematics,* 12, 428-454.

Binmore, K., Swierzbinski, J. and C. Proulx (2001), "Does Minimax Work? An Experimental Study," *Economic Journal* 111 (473), 445-464.

Brown, J. and Rosenthal, R. (1990), "Testing the Minimax Hypothesis: A Reexamination of O'Neill's Experiment," *Econometrica,* 58, 1065-1081.

Camerer, C. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press.

Camerer, C. (1995), "Individual Decision Making," in *Handbook of Experimental Economics*, edited by J. H. Hagel and A. E. Roth. Princeton University Press, 587-703.

Camerer, C.; Loewenstein, G. and D. Prelec (2005), "Neuroeconomics: How Neuroscience Can Inform Economics," *Journal of Economic Literature* 43, 9-64.

Chiappori, P. A., Levitt, S. D. and Groseclose, T. (2002): "Testing Mixed Strategy Equilibrium When Players Are Heterogeneous: The Case of Penalty Kicks," *American Economic Review* 92(4), 1138-1151.

Erev, I. and Roth, A. E. (1998), "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review,* 88, 848-881.

Fédération Internationale de Football Association (2005), *Official Laws of the Game.* Triumph Books, Chicago, IL.

Gibbons, J. and Chakraborti, S. (1992), *Nonparametric Statistical Inference*, New York: Marcel Dekker.

Giggerenzer, G. and P. M. Todd (1999), *Simple Heuristics That Make Us Smart*, Oxford University Press, New York.

Gilboa, I. and D. Schmeidler (2001), *A Theory of Case-Based Decision Theory,* Cambridge University Press, Cambridge, UK.

Gneezy, U.; A. Rustichini and A. Vostruknutov (2005), "Backward Induction as a Cognitive Process," Mimeo, University of Minnesota.

Grether, D. M. and C. R. Plott (1984), "The Effects of Market Practices in Oligopolistic Markets: An Experimental Examination of the Ethyl Case," *Economic Inquiry,* 22, 479-507.

Haier, R. J.; B. V. Siegel, Jr.; A. MacLachlan, E. Soderling, S. Lottenberg, and M. S. Buchsbaum (1992), "Regional Glucose Metabolic Changes After Learning a Complex Visuospatial/Motor Task: a PET Study," *Brain Research* 570 (1-2): 134-143.

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

Hong, J. T. and C. R. Plott (1982), "Rate Filing Policies for Inland Water Transportation: An Experimental Approach," *Bell Journal of Economics* 13, pp. 1-19.

Hume, D. (1749), *An Enquiry Concerning Human Understanding,* Open Court Publishing, Chicago and La Salle, 9th Edition.

Kagel, J. H. and A. E. Roth (1995), *Handbook of Experimental Economics*, Princeton University Press.

Lazear, E.; U. Malmendier, and R. A. Weber (2005), "Sorting in Experiments," Mimeo, Stanford University.

Lo, A. W. and D. V. Repin (2002), "The Psychophysiology of Real-Time Financial Risk Processing," *Journal of Cognitive Neuroscience* 14(3), 323-339.

Marca (2005), *Guia de la Liga 2006.* Recoletos Grupo de Comunicación, Madrid.

Miller, C. (1998), *He Always Puts It to the Right. A Concise History of the Penalty Kick.* Victor Gollancz, St. Edmundsbury Press Ltd., London.

Mookherjee, D. and Sopher, B. (1994), "Learning Behavior in an Experimental Matching Pennies Game," *Games and Economic Behavior,* 7, 62-91.

Neuringer, A. J. (1986), "Can People Behave "Randomly"?: The Role of Feedback," *Journal of Experimental Psychology: General,* 115, 62-75.

Neuringer, A. J. (2002), "Operant Variability: Evidence, Function, and Theory," *Psychonomic Bulletin and Review,* 9, 672-705.

O'Neill, B. (1987), "Nonmetric Test of the Minimax Theory of Two-Person Zero-Sum Games," *Proceedings of the National Academy of Sciences,* 84, 2106-2109.

Palacios-Huerta, I. (2003), "Professionals Play Minimax," *Review of Economic Studies* 70 (2), 395-415.

Palacios-Huerta, I. and O. Volij (2006), "Appendix to Experientia Docet: Professionals Play Minimax in Laboratory Experiments," Mimeo, Brown University.

Rapoport, A., and Boebel, R. B. (1992), "Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis," *Games and Economic Behavior,* 4, 261-283.

Rapoport, A., and Budescu, D. V. (1992), "Generation of Random Sequences in Two-Person Strictly Competitive Games," *Journal of Experimental Psychology: General,* 121, 352-363.

Rubinstein, A. (2005), "Instinctive and Cognitive Reasoning: Response Times Study," Mimeo, New York University.

Shachat, J. M. (2002), "Mixed Strategy Play and the Minimax Hypothesis," *Journal of Economic Theory,* 104, 189-226.

Selten, R. (1998), "Features of Experimentally Observed Bounded Rationality," *European Economic Review*, 42, 2-5, 413-436.

Simon, H. A. (1983), *Reason in Human Affairs*, Stanford University Press.

Slonim, R. E.; A. Roth and I. Erev (2003), "Regularities in Play of Zero-Sum Games with Unique Mixed-Strategy Equilibria," Mimeo, Case Western Reserve University.

Smith, V. L. (2005), "Behavioral Economics Research and the Foundations of Economics," *Journal of Socio-Economics,* 34, 135-150.

Tversky, A. and D. Kanheman (1971), "The Belief in the Law of Small Numbers," *Psychological Bulletin,* 76, 105-110.

von Neumann, J. (1928), "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen,* 100, 295-320.

Weibull, J. (2004), "Testing Game Theory," Mimeo, Department of Economics, Boston University.

Walker, M. and Wooders, J. (2001), "Minimax Play at Wimbledon," *American Economic Review*, 91, 1521-1538.

Table A1

Relative Frequencies of Choices and Win Percentages
in Penalty Kick's Experiment for Professional Players

A. Frequencies

		Column Player Choice: L	Column Player Choice: R	Marginal Frequencies for Row Player:
Row Player Choice	L	0.152 (0.165) [0.0068]	0.182 (0.198) [0.0073]	0.333 (0.364) [0.0088]
	R	0.310 (0.289) [0.0083]	0.356 (0.347) [0.0087]	0.667 (0.636) [0.0088]
Marginal Frequencies for Column Player:		0.462 (0.455) [0.009]	0.538 (0.545) [0.009]	

B. Win Percentages

Observed Row Player Win Percentage:	0.7947
Minimax Row Player Win Percentage:	0.7909
Minimax Row Player Win Std. Deviation:	0.0074

Notes: In Panel A numbers in parentheses represent Minimax predicted relative frequencies. Numbers in brackets represent standard deviations for observed relative frequencies under the Minimax hypothesis. In Panel B, Minimax Row Player Win Percentage and Std. Deviation are the mean and the std. deviation of the observed row player mean percentage win under the Minimax hypothesis.

Table A2

Marginal Frequencies and Action Pair Frequencies in Penalty Kick's Experiment for Professional Players

Pair #	Marginal Frequencies Row L	Marginal Frequencies Column L	Pair Frequencies LL	LR	RL	RR	χ^2 *p*-value
1	0.320	0.453	0.140	0.180	0.313	0.367	0.729
2	0.360	0.380*	0.127	0.233	0.253	0.387	0.305
3	0.307	0.427	0.127	0.180	0.300	0.393	0.459
4	0.327	0.460	0.153	0.173	0.307	0.367	0.819
5	0.327	0.493	0.153	0.173	0.340	0.333	0.568
6	0.340	0.480	0.140	0.200	0.340	0.320	0.525
7	0.287**	0.427	0.133	0.153	0.293	0.420	0.190
8	0.320	0.460	0.100	0.220	0.360	0.320	0.068*
9	0.307	0.467	0.133	0.173	0.333	0.360	0.479
10	0.313	0.480	0.167	0.147	0.313	0.373	0.454
11	0.353	0.480	0.180	0.173	0.300	0.347	0.866
12	0.427*	0.480	0.193	0.233	0.287	0.287	0.359
13	0.367	0.473	0.167	0.200	0.307	0.327	0.952
14	0.327	0.447	0.153	0.173	0.293	0.380	0.782
15	0.340	0.553**	0.173	0.167	0.380	0.280	0.071*
16	0.320	0.473	0.160	0.160	0.313	0.367	0.659
17	0.347	0.467	0.200	0.147	0.267	0.387	0.256
18	0.327	0.440	0.140	0.187	0.300	0.373	0.791
19	0.327	0.440	0.140	0.187	0.300	0.373	0.791
20	0.327	0.460	0.153	0.173	0.307	0.367	0.819

Notes: ** and * denote rejections at the 5 and 10 percent levels respectively of the Minimax binomial model for the marginal frequencies of the row and column players. In the last column they denote rejections of the joint hypothesis that both players in a pair choose actions with the equilibrium frequency.

Table A3

Observed and Expected Win Percentages under the Independence Hypothesis
for Professional Players

Pair #	Observed Row Win %	Expected Row Win %	Mixture Effect	Correlated Play Effect
1	0.8000	0.7909	0.0000	0.0091
2	0.8000	0.7908	-0.0001	0.0092
3	0.7867	0.7900	-0.0009	-0.0034
4	0.7867	0.7910	0.0001	-0.0044
5	0.8000	0.7917	0.0008	0.0083
6	0.7933	0.7912	0.0003	0.0021
7	0.8067	0.7897	-0.0012	0.0169
8	0.7800	0.7910	0.0001	-0.0110
9	0.7933	0.7913	0.0004	0.0020
10	0.8067	0.7916	0.0007	0.0151
11	0.7733	0.7911	0.0001	-0.0177
12	0.8200	0.7900	-0.0009	0.0300
13	0.7667	0.7909	0.0000	-0.0242
14	0.8000	0.7907	-0.0002	0.0093
15	0.7933	0.7922	0.0013	0.0011
16	0.8067	0.7914	0.0005	0.0153
17	0.7933	0.7910	0.0001	0.0023
18	0.7867	0.7906	-0.0003	-0.0039
19	0.7867	0.7906	-0.0003	-0.0039
20	0.8133	0.7910	0.0001	0.0223
Mean Absolute Value:			0.0004	0.0106
Chi-Square Statistic:			0.0059	3.2344
Chi-Square Probability Value:			0.9999	0.9999

Notes: Expected winning percentages assume independent play with mixtures observed over all 150 games for each pair. Mixture effects are defined as the difference between the values in column 2 and 0.7909. Correlation effects are defined as the difference between columns 1 and 2.

Table A4 - Testing that Professional Players Equate their Strategies' Payoffs to the Equilibrium Rates

		L		R		Pearson	
Pair #	**Player**	**Success**	**Fail**	**Success**	**Fail**	**statistic**	***p*-value**
1	Row	0.260	0.060	0.540	0.140	1.360	0.715
	Column	0.080	0.373	0.120	0.427	0.491	0.921
2	Row	0.300	0.060	0.500	0.140	0.645	0.886
	Column	0.047	0.333	0.153	0.467	6.441	0.092*
3	Row	0.233	0.073	0.553	0.140	2.351	0.503
	Column	0.100	0.327	0.113	0.460	0.774	0.856
4	Row	0.247	0.080	0.540	0.133	1.306	0.728
	Column	0.107	0.353	0.107	0.433	0.302	0.960
5	Row	0.280	0.047	0.520	0.153	2.278	0.517
	Column	0.100	0.393	0.100	0.407	0.989	0.804
6	Row	0.280	0.060	0.513	0.147	0.776	0.855
	Column	0.080	0.400	0.127	0.393	1.755	0.625
7	Row	0.207	0.080	0.600	0.113	6.673	0.083*
	Column	0.093	0.333	0.100	0.473	1.161	0.762
8	Row	0.273	0.047	0.507	0.173	3.640	0.303
	Column	0.113	0.347	0.107	0.433	0.670	0.880
9	Row	0.233	0.073	0.560	0.133	2.508	0.474
	Column	0.113	0.353	0.093	0.440	1.134	0.769
10	Row	0.247	0.067	0.560	0.127	2.051	0.562
	Column	0.093	0.387	0.100	0.420	0.617	0.892
11	Row	0.260	0.093	0.513	0.133	1.018	0.797
	Column	0.107	0.373	0.120	0.400	0.683	0.877
12	Row	0.327	0.100	0.493	0.080	5.132	0.162
	Column	0.073	0.407	0.107	0.413	1.857	0.603
13	Row	0.287	0.080	0.480	0.153	0.657	0.883
	Column	0.100	0.373	0.133	0.393	1.112	0.774
14	Row	0.247	0.080	0.553	0.120	1.843	0.606
	Column	0.080	0.367	0.120	0.433	0.426	0.935
15	Row	0.260	0.080	0.533	0.127	0.743	0.863
	Column	0.093	0.460	0.113	0.333	7.563	0.056*
16	Row	0.253	0.067	0.553	0.127	1.578	0.664
	Column	0.073	0.400	0.120	0.407	1.687	0.640
17	Row	0.253	0.093	0.540	0.113	2.043	0.564
	Column	0.120	0.347	0.087	0.447	2.119	0.548
18	Row	0.253	0.073	0.533	0.140	0.950	0.813
	Column	0.087	0.353	0.127	0.433	0.337	0.953
19	Row	0.260	0.067	0.527	0.147	0.942	0.815
	Column	0.073	0.367	0.140	0.420	1.696	0.638
20	Row	0.260	0.067	0.553	0.120	1.509	0.680
	Column	0.093	0.367	0.093	0.447	0.671	0.880

Notes: ** and * denote rejections at the 5 and 10 percent levels respectively.

Table A5

Runs Tests in Penalty Kick's Experiment
for Professional Players

		Choices		Runs		
Pair	Player	R	L	r^j	$F(r^j\text{-}1)$	$F(r^j)$
1	Row	102	48	72	0.840	0.877
	Column	82	68	69	0.129	0.167
2	Row	96	54	74	0.727	0.779
	Column	93	57	72	0.488	0.554
3	Row	104	46	64	0.404	0.469
	Column	86	64	82	0.884	0.913
4	Row	101	49	69	0.604	0.682
	Column	81	69	75	0.433	0.499
5	Row	101	49	79	0.985**	0.992
	Column	76	74	80	0.717	0.770
6	Row	99	51	74	0.830	0.869
	Column	78	72	89	0.981**	0.987
7	Row	107	43	53	0.025	0.041*
	Column	86	64	72	0.315	0.375
8	Row	102	48	69	0.655	0.730
	Column	81	69	69	0.124	0.160
9	Row	104	46	63	0.323	0.404
	Column	80	70	67	0.066	0.089
10	Row	103	47	58	0.065	0.089
	Column	78	72	85	0.922	0.943
11	Row	97	53	66	0.235	0.289
	Column	78	72	69	0.113	0.147
12	Row	86	64	68	0.125	0.162
	Column	78	72	77	0.541	0.605
13	Row	95	55	71	0.484	0.559
	Column	79	71	80	0.729	0.781
14	Row	101	49	72	0.802	0.845
	Column	83	67	63	0.018	0.027*
15	Row	99	51	68	0.441	0.507
	Column	67	83	68	0.103	0.135
16	Row	102	48	67	0.509	0.592
	Column	79	71	74	0.353	0.416
17	Row	98	52	71	0.605	0.679
	Column	80	70	72	0.246	0.301
18	Row	101	49	62	0.156	0.199
	Column	84	66	71	0.231	0.285
19	Row	101	49	68	0.539	0.604
	Column	84	66	78	0.666	0.724
20	Row	101	49	75	0.918	0.947
	Column	81	69	71	0.204	0.254

Notes: ** and * denote rejections at the 5 and 10 percent levels respectively.

Table A6

Results of Significance Tests from Logit Equations for the Choice of Choice of Right (R)
for Professional Players

Estimating Equation: $R = G\left[\beta_1 + \beta_2 \,\mathrm{lag}(R) + \beta_3 \,\mathrm{lag}(R)\,\mathrm{lag}(W) + \beta_4 \,\mathrm{lag}(L)\,\mathrm{lag}(W) + \beta_5 \,\mathrm{lag}(R^*) + \beta_6 R^*\right]$

Null Hypothesis	Player Pairs Whose Behavior Allows Rejection of the Null Hypothesis at the:	5 percent level	10 percent level
(1). $\beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$	Row:	-	11
	Column:	3,8	3,6,8
(2). $\beta_2 = 0$	Row:	5	5,7
	Column:	5	3,5,11
(3). $\beta_3 = 0$	Row:	-	-
	Column:	-	5
(4). $\beta_4 = 0$	Row:	11	11
	Column:	3	3,9
(5). $\beta_5 = 0$	Row:	17	17
	Column:	8,17	8,6,17
(6). $\beta_6 = 0$	Row:	8	8
	Column:	-	-

Notes: The symbols R and R^* denote the choice of "right" by a player and by his opponent respectively. The symbol W denotes "win" by a player. The term "lag" refers to the previous choice or outcome in the ordered sequence. The function G[x] denotes the function exp(x)/[1+exp(x)]. Rejections are based on likelihood-ratio tests.

Table B1

Relative Frequencies of Choices and Win Percentages
in Penalty Kick's Experiment for College Students

A. Frequencies

		Column Player Choice: L	Column Player Choice: R	Marginal Frequencies for Row Player:
Row Player Choice	**L**	0.168 (0.165) [0.0068]	0.233 (0.198) [0.0073]	0.401 (0.364) [0.0088]
	R	0.228 (0.289) [0.0083]	0.370 (0.347) [0.0087]	0.599 (0.636) [0.0088]
Marginal Frequencies for Column Player:		0.397 (0.455) [0.009]	0.603 (0.545) [0.009]	

B. Win Percentages

Observed Row Player Win Percentage:	0.7877
Minimax Row Player Win Percentage:	0.7909
Minimax Row Player Win Std. Deviation:	0.0074

Notes: Numbers in parentheses represent Minimax predicted relative frequencies. Numbers in brackets represent standard deviations for observed relative frequencies under the Minimax hypothesis. In Panel B, Minimax Row Player Win Percentage and Std. Deviation are the mean and the std. deviation of the observed row player mean percentage win under the Minimax hypothesis.

Table B2

Marginal Frequencies and Action Pair Frequencies in Penalty Kick's Experiment for College Students

	Marginal Frequencies		Pair Frequencies				
Pair #	**Row L**	**Column L**	**LL**	**LR**	**RL**	**RR**	χ^2 ***p*-value**
1	0.360	0.387*	0.147	0.213	0.240	0.400	0.399
2	0.427*	0.387*	0.160	0.267	0.227	0.347	0.134
3	0.427*	0.387*	0.160	0.267	0.227	0.347	0.134
4	0.427*	0.433	0.173	0.253	0.260	0.313	0.350
5	0.413	0.387*	0.167	0.247	0.220	0.367	0.220
6	0.413	0.387*	0.147	0.267	0.240	0.347	0.164
7	0.427*	0.407	0.207	0.220	0.200	0.373	0.096*
8	0.407	0.387*	0.140	0.267	0.247	0.347	0.168
9	0.427*	0.393	0.187	0.240	0.207	0.367	0.143
10	0.380	0.367**	0.133	0.247	0.233	0.387	0.172
11	0.427*	0.480	0.167	0.260	0.313	0.260	0.091*
12	0.420	0.400	0.213	0.207	0.187	0.393	0.036**
13	0.427*	0.393	0.233	0.193	0.160	0.413	0.002**
14	0.287**	0.460	0.140	0.147	0.320	0.393	0.260
15	0.220**	0.440	0.100	0.120	0.340	0.440	0.004**
16	0.460**	0.300**	0.120	0.340	0.180	0.360	0.000**
17	0.427*	0.367**	0.160	0.267	0.207	0.367	0.064*
18	0.407	0.387*	0.153	0.253	0.233	0.360	0.250
19	0.427*	0.393	0.233	0.193	0.160	0.413	0.002**
20	0.420	0.393	0.227	0.193	0.167	0.413	0.004**

Notes: ** and * denote rejections at the 5 and 10 percent levels respectively of the Minimax binomial model for the marginal frequencies of the row and column players. In the last column they denote rejections of the joint hypothesis that both players in a pair choose actions with the equilibrium frequency.

Table B3

Observed and Expected Win Percentages under the Independence Hypothesis for College Students

Pair #	Observed Row Win %	Expected Row Win %	Mixture Effect	Correlated Play Effect
1	0.8333	0.7908	-0.0001	0.0426
2	0.7867	0.7933	0.0024	-0.0066
3	0.7867	0.7933	0.0024	-0.0066
4	0.7867	0.7916	0.0007	-0.0050
5	0.7333	0.7928	0.0019	-0.0594
6	0.8200	0.7928	0.0019	0.0272
7	0.7733	0.7926	0.0017	-0.0192
8	0.7933	0.7925	0.0016	0.0008
9	0.7733	0.7930	0.0021	-0.0197
10	0.8000	0.7917	0.0008	0.0083
11	0.8467	0.7900	-0.0009	0.0566
12	0.7267	0.7926	0.0017	-0.0659
13	0.7267	0.7930	0.0021	-0.0664
14	0.8067	0.7911	0.0002	0.0155
15	0.8133	0.7898	-0.0011	0.0236
16	0.8200	0.7991	0.0082	0.0209
17	0.7733	0.7940	0.0030	-0.0206
18	0.8267	0.7925	0.0016	0.0342
19	0.7867	0.7930	0.0021	-0.0064
20	0.7400	0.7928	0.0019	-0.0528
Mean Absolute Value:			0.0019	0.0279
Chi-Square Statistic:			0.1139	22.3481
Chi-Square Probability Value:			0.9999	0.3219

Notes: Expected winning percentages assume independent play with mixtures observed over all 150 games for each pair. Mixture effects are defined as the difference between the values in column 2 and 0.7909. Correlation effects are defined as the difference between columns 1 and 2.

Table B4 - Testing that College Students Equate their Strategies' Payoffs to the Equilibrium Rates

Pair #	Player	L Success	L Fail	R Success	R Fail	Pearson statistic	*p*-value
1	Row	0.313	0.047	0.520	0.120	2.322	0.508
	Column	0.053	0.333	0.113	0.500	4.668	0.198
2	Row	0.360	0.067	0.427	0.147	4.866	0.182
	Column	0.107	0.280	0.107	0.507	4.892	0.180
3	Row	0.353	0.073	0.433	0.140	3.781	0.286
	Column	0.060	0.327	0.153	0.460	4.702	0.195
4	Row	0.360	0.067	0.427	0.147	4.866	0.182
	Column	0.093	0.340	0.120	0.447	0.291	0.962
5	Row	0.293	0.120	0.440	0.147	5.234	0.155
	Column	0.120	0.267	0.147	0.467	6.411	0.093*
6	Row	0.367	0.047	0.453	0.133	5.706	0.127
	Column	0.067	0.320	0.113	0.500	3.559	0.313
7	Row	0.327	0.100	0.447	0.127	2.931	0.402
	Column	0.107	0.300	0.120	0.473	2.348	0.503
8	Row	0.347	0.060	0.447	0.147	3.491	0.322
	Column	0.053	0.333	0.153	0.460	5.345	0.148
9	Row	0.340	0.087	0.433	0.140	3.168	0.366
	Column	0.120	0.273	0.107	0.500	5.789	0.122
10	Row	0.307	0.073	0.493	0.127	0.280	0.964
	Column	0.053	0.313	0.147	0.487	6.096	0.107
11	Row	0.387	0.040	0.460	0.113	8.677	0.034**
	Column	0.060	0.420	0.093	0.427	4.037	0.257
12	Row	0.300	0.120	0.427	0.153	6.108	0.106
	Column	0.140	0.260	0.133	0.467	8.243	0.041**
13	Row	0.293	0.133	0.433	0.140	8.008	0.046**
	Column	0.147	0.247	0.127	0.480	10.549	0.014**
14	Row	0.207	0.080	0.600	0.113	6.673	0.083*
	Column	0.093	0.367	0.100	0.440	0.311	0.958
15	Row	0.173	0.047	0.640	0.140	14.135	0.003**
	Column	0.047	0.393	0.140	0.420	5.102	0.164
16	Row	0.373	0.087	0.447	0.093	6.791	0.079*
	Column	0.060	0.240	0.120	0.580	15.620	0.001**
17	Row	0.320	0.107	0.453	0.120	3.335	0.343
	Column	0.120	0.247	0.107	0.527	9.523	0.023**
18	Row	0.367	0.040	0.460	0.133	6.381	0.094*
	Column	0.060	0.327	0.113	0.500	4.025	0.259
19	Row	0.347	0.080	0.440	0.133	3.045	0.385
	Column	0.093	0.300	0.120	0.487	2.590	0.459
20	Row	0.280	0.140	0.460	0.120	8.854	0.031**
	Column	0.140	0.253	0.120	0.487	9.002	0.029**

Notes: ** and * denote rejections at the 5 and 10 percent levels respectively.

Table B5

Runs Tests in Penalty Kick's Experiment
for College Students

Pair	Player	Choices R	Choices L	Runs r^j	$F(r^j\text{-}1)$	$F(r^j)$
1	Row	96	54	69	0.383	0.457
	Column	92	58	61	0.022	0.033*
2	Row	86	64	90	0.995**	0.997
	Column	92	58	70	0.324	0.386
3	Row	86	64	65	0.049	0.069
	Column	92	58	91	0.999**	1.000
4	Row	86	64	77	0.637	0.699
	Column	85	65	82	0.873	0.904
5	Row	88	62	78	0.737	0.788
	Column	92	58	78	0.823	0.863
6	Row	88	62	72	0.352	0.415
	Column	92	58	71	0.386	0.456
7	Row	86	64	65	0.049	0.069
	Column	89	61	66	0.091	0.121
8	Row	89	61	84	0.958*	0.971
	Column	92	58	58	0.006	0.009**
9	Row	86	64	79	0.754	0.804
	Column	91	59	80	0.883	0.913
10	Row	93	57	82	0.958*	0.970
	Column	95	55	66	0.182	0.229
11	Row	86	64	76	0.574	0.637
	Column	78	72	69	0.113	0.147
12	Row	87	63	63	0.026	0.038*
	Column	90	60	85	0.976**	0.984
13	Row	86	64	68	0.125	0.162
	Column	91	59	88	0.995**	0.997
14	Row	107	43	82	0.999**	0.999
	Column	81	69	66	0.049	0.068
15	Row	117	33	67	0.999**	0.999
	Column	84	66	82	0.863	0.896
16	Row	81	69	73	0.309	0.369
	Column	105	45	70	0.863	0.896
17	Row	86	64	74	0.441	0.507
	Column	95	55	69	0.348	0.419
18	Row	89	61	84	0.958*	0.971
	Column	92	58	83	0.963*	0.976
19	Row	86	64	76	0.574	0.637
	Column	91	59	76	0.692	0.747
20	Row	87	63	72	0.332	0.394
	Column	91	59	81	0.913	0.938

Notes: ** and * denote rejections at the 5 and 10 percent levels respectively.

Table B6

Results of Significance Tests from Logit Equations for the Choice of Choice of Right (*R*)
for College Students

Estimating Equation: $R = G\left[\beta_1 + \beta_2\,\mathrm{lag}(R) + \beta_3\,\mathrm{lag}(R)\,\mathrm{lag}(W) + \beta_4\,\mathrm{lag}(L)\,\mathrm{lag}(W) + \beta_5\,\mathrm{lag}(R^*) + \beta_6\,R^*\right]$

Null Hypothesis	Player Pairs Whose Behavior Allows Rejection of the Null Hypothesis at the:		
		5 percent level	10 percent level
(1). $\beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$	Row:	11,12,13,14,15,19,20	2,11,12,13,14,15,19,20
	Column:	3,9,12,13,20	3,9,12,13,19,20
(2). $\beta_2 = 0$	Row:	2,3,10,14,15	2,3,7,10,14,15
	Column:	13	13,20
(3). $\beta_3 = 0$	Row:	12	12
	Column:	-	-
(4). $\beta_4 = 0$	Row:	-	6
	Column:	9,13	4,9,13,20
(5). $\beta_5 = 0$	Row:	11,13	1,9,11,13
	Column:	17	11,17,18
(6). $\beta_6 = 0$	Row:	12,13,19,20	7,11,12,13,19,20
	Column:	12,13	7,11,12,13,19,20

Notes: The symbols *R* and *R** denote the choice of “right” by a player and by his opponent respectively. The symbol *W* denotes “win.” The term “lag” refers to the previous choice or outcome in the ordered sequence. The function G[x] denotes the function exp(x)/[1+exp(x)]. Rejections are based on likelihood-ratio tests.

Table C1

Relative Frequencies of Card Choices in O'Neill's Experiment
Professional Players

A. Frequencies

Row Player Choice	Column Player Choice: 1	2	3	*J*	Marginal Frequencies for Row Player:
1	0.037 (0.040) [0.003]	0.042 (0.040) [0.003]	0.039 (0.040) [0.003]	0.083 (0.080) [0.004]	0.201 (0.200) [0.006]
2	0.042 (0.040) [0.003]	0.038 (0.040) [0.003]	0.044 (0.040) [0.003]	0.079 (0.080) [0.004]	0.203 (0.200) [0.006]
3	0.038 (0.040) [0.003]	0.037 (0.040) [0.003]	0.040 (0.040) [0.003]	0.083 (0.080) [0.004]	0.198 (0.200) [0.006]
J	0.084 (0.080) [0.004]	0.082 (0.080) [0.004]	0.081 (0.080) [0.004]	0.153 (0.160) [0.006]	0.398 (0.400) [0.008]
Marginal Frequencies For Column Player:	0.200 (0.200) [0.006]	0.198 (0.200) [0.006]	0.204 (0.200) [0.006]	0.398 (0.400) [0.008]	

B. Win Percentages

Observed Row Player Win Percentage: 0.3945
Minimax Row Player Win Percentage: 0.4000
Minimax Row Player Win Std. Deviation: 0.0077

Notes: In Panel A, numbers in parentheses represent Minimax predicted relative frequencies, and numbers in brackets represent standard deviations for observed relative frequencies under the Minimax hypothesis. In Panel B, Minimax Row Player Win Percentage and Std. Deviation are the mean and the std. deviation of the observed row player mean percentage win under the Minimax hypothesis.

Table C2

Relative Frequencies of Card Choices in O'Neill's Experiment by Player Pair
Professional Players

Pair #	Row Player (R) Choice				Column Player (C) Choice				Rejection of Minimax models	
	1	**2**	**3**	***J***	**1**	**2**	**3**	***J***	**at 5%:**	**at 10%:**
1	0.190	0.225	0.290	0.295	0.195	0.185	0.210	0.410	a,c,R3,RJ	a,c,R3,RJ
2	0.205	0.215	0.245	0.335	0.200	0.205	0.250	0.345	RJ	R3,RJ,C3,CJ
3	0.210	0.195	0.200	0.395	0.195	0.175	0.205	0.425		
4	0.215	0.205	0.180	0.400	0.145	0.185	0.225	0.445	C1	C1
5	0.180	0.195	0.205	0.420	0.200	0.195	0.210	0.395		
6	0.210	0.205	0.185	0.400	0.205	0.185	0.205	0.405		
7	0.215	0.215	0.130	0.440	0.205	0.190	0.205	0.400	R3	R3
8	0.195	0.215	0.195	0.395	0.225	0.150	0.205	0.420		C2
9	0.185	0.195	0.215	0.405	0.205	0.180	0.205	0.410		
10	0.175	0.180	0.170	0.475	0.195	0.195	0.215	0.395	RJ	RJ
11	0.205	0.190	0.170	0.435	0.250	0.200	0.205	0.345		C1,CJ
12	0.200	0.200	0.195	0.405	0.195	0.200	0.205	0.400		
13	0.215	0.185	0.195	0.405	0.195	0.215	0.190	0.400		
14	0.185	0.185	0.205	0.425	0.205	0.290	0.195	0.310	b,c,C2,CJ	b,c,C2,CJ
15	0.215	0.200	0.170	0.415	0.210	0.185	0.200	0.405		
16	0.205	0.195	0.195	0.405	0.195	0.165	0.175	0.465		CJ
17	0.205	0.230	0.190	0.375	0.225	0.215	0.205	0.355		
18	0.210	0.195	0.180	0.415	0.205	0.245	0.210	0.340		C2,CJ
19	0.205	0.220	0.235	0.340	0.170	0.205	0.175	0.450		RJ
20	0.195	0.210	0.200	0.395	0.185	0.205	0.180	0.430		

Notes: The pairs of capital letters denote rejection of the Minimax binomial model for a given card (1,2,3,J) for a player (R,C).
[a] denotes rejection of Minimax multinomial model for all cards chosen by the row player based on Pearson statistic and $\chi^2(3)$.
[b] denotes rejection of Minimax multinomial model for all cards chosen by the column player based on Pearson statistic and $\chi^2(3)$.
[c] denotes rejection of Minimax multinomial model for all cards chosen by both players based on Pearson statistic and $\chi^2(6)$.

Table C3

Observed and Expected Win Percentages under the Independence Hypothesis

Pair #	Observed Row Win %	Expected Row Win %	Mixture Effect	Correlated Play Effect
1	0.3400	0.3973	-0.0027	-0.0573
2	0.3450	0.4048	0.0048	-0.0598
3	0.3950	0.3997	-0.0003	-0.0047
4	0.4150	0.4014	0.0014	0.0136
5	0.4000	0.3997	-0.0003	0.0003
6	0.4000	0.4001	0.0001	-0.0001
7	0.4000	0.4004	0.0004	-0.0004
8	0.4050	0.4007	0.0007	0.0043
9	0.4000	0.4000	0.0000	0.0000
10	0.3900	0.3995	-0.0005	-0.0095
11	0.4100	0.3961	-0.0040	0.0140
12	0.4050	0.4000	0.0000	0.0050
13	0.4000	0.4003	0.0003	-0.0003
14	0.4200	0.3970	-0.0031	0.0231
15	0.3900	0.4000	0.0000	-0.0100
16	0.3950	0.4004	0.0004	-0.0054
17	0.3900	0.4017	0.0017	-0.0117
18	0.4050	0.3986	-0.0014	0.0064
19	0.3950	0.3949	-0.0051	0.0001
20	0.3900	0.3996	-0.0004	-0.0096
Mean Absolute Value:			0.0014	0.0118
Chi-Square Statistic:			0.0747	6.9422
Chi-Square Probability Value:			0.9999	0.9969

Notes: Expected winning percentages assume independent play with mixtures observed over all 200 games for each pair. Mixture effects are defined as the difference between the values in column 2 and 0.40. Correlation effects are defined as the difference between columns 1 and 2.

Table C4 - Testing that Professional Players Equate their Strategies' Payoffs to the Equilibrium Rates

		Mixtures		Win Rates			
Pair #	**Player**	**Joker**	**Non-Joker**	**Joker**	**Non-Joker**	**Pearson**	***p*-value**
1	R	0.295	0.705	0.407	0.312	14.535	0.002** ‡
	C	0.410	0.590	0.707	0.627	4.472	0.215
2	R	0.335	0.665	0.403	0.316	7.878	0.049** ‡
	C	0.345	0.655	0.609	0.679	6.295	0.098*
3	R	0.395	0.605	0.367	0.413	0.462	0.927
	C	0.425	0.575	0.659	0.565	2.378	0.498
4	R	0.400	0.600	0.388	0.433	0.608	0.895
	C	0.445	0.555	0.652	0.532	4.795	0.187 †
5	R	0.420	0.580	0.429	0.379	0.833	0.841
	C	0.395	0.605	0.544	0.636	1.701	0.637
6	R	0.400	0.600	0.388	0.408	0.087	0.993
	C	0.405	0.595	0.617	0.588	0.191	0.979
7	R	0.440	0.560	0.352	0.438	2.865	0.413
	C	0.400	0.600	0.613	0.592	0.087	0.993
8	R	0.395	0.605	0.456	0.372	1.431	0.698
	C	0.420	0.580	0.571	0.612	0.701	0.873
9	R	0.405	0.595	0.358	0.429	1.024	0.795
	C	0.410	0.590	0.646	0.568	1.337	0.720
10	R	0.475	0.525	0.358	0.419	5.660	0.129
	C	0.395	0.605	0.570	0.636	0.993	0.803
11	R	0.435	0.565	0.368	0.442	2.229	0.526
	C	0.345	0.655	0.536	0.618	3.729	0.292
12	R	0.405	0.595	0.383	0.420	0.323	0.956
	C	0.400	0.600	0.613	0.583	0.191	0.979
13	R	0.405	0.595	0.420	0.387	0.243	0.970
	C	0.400	0.600	0.575	0.617	0.347	0.951
14	R	0.425	0.575	0.353	0.470	3.576	0.311
	C	0.310	0.690	0.516	0.609	8.208	0.042** ‡
15	R	0.415	0.585	0.373	0.402	0.441	0.932
	C	0.405	0.595	0.617	0.605	0.135	0.987
16	R	0.405	0.595	0.420	0.378	0.389	0.943
	C	0.465	0.535	0.634	0.579	4.222	0.238
17	R	0.375	0.625	0.387	0.392	0.608	0.895
	C	0.355	0.645	0.592	0.620	1.941	0.585
18	R	0.415	0.585	0.301	0.479	6.628	0.085*
	C	0.340	0.660	0.632	0.576	3.608	0.307
19	R	0.340	0.660	0.412	0.386	3.146	0.370
	C	0.450	0.550	0.689	0.536	7.118	0.068*
20	R	0.395	0.605	0.367	0.405	0.385	0.943
	C	0.430	0.570	0.663	0.570	2.670	0.445

Notes: ** and * denote rejections at the 5 and 10 percent levels respectively. ‡ and † denote the players for whom rejections at the 5 and 10 percent levels of the tests of equality of winning probabilities to the equilibrium rate are also found when the four cards are treated individually.

Table C5

Runs Tests in O'Neill's Experiment
Professional Players

Pair	Player	Choices: Joker	Choices: Non-Joker	Runs r^j	$F(r^j\text{-}1)$	$F(r^j)$
1	R	59	141	90	0.821	0.856
	C	82	118	88	0.067	0.087
2	R	67	133	94	0.707	0.754
	C	69	131	92	0.508	0.564
3	R	79	121	90	0.147	0.182
	C	85	115	100	0.543	0.599
4	R	80	120	98	0.530	0.586
	C	89	111	115	0.983**	0.988
5	R	84	116	94	0.236	0.283
	C	79	121	96	0.436	0.493
6	R	80	120	110	0.968*	0.977
	C	81	119	95	0.334	0.391
7	R	88	112	89	0.056	0.074
	C	80	120	100	0.644	0.696
8	R	79	121	94	0.324	0.377
	C	84	116	102	0.672	0.722
9	R	81	119	103	0.773	0.816
	C	82	118	91	0.143	0.180
10	R	95	105	98	0.322	0.375
	C	79	121	98	0.554	0.610
11	R	87	113	107	0.850	0.882
	C	69	131	97	0.786	0.833
12	R	81	119	91	0.155	0.194
	C	80	120	100	0.644	0.696
13	R	81	119	93	0.235	0.284
	C	80	120	93	0.252	0.303
14	R	85	115	89	0.068	0.090
	C	62	138	87	0.488	0.563
15	R	83	117	101	0.635	0.690
	C	81	119	99	0.563	0.622
16	R	81	119	114	0.992**	0.994
	C	93	107	108	0.840	0.873
17	R	75	125	97	0.601	0.662
	C	71	129	101	0.889	0.918
18	R	83	117	98	0.465	0.521
	C	68	132	78	0.019	0.027*
19	R	68	132	90	0.422	0.478
	C	90	110	96	0.260	0.308
20	R	79	121	96	0.436	0.493
	C	86	114	90	0.084	0.108

Notes: ** and * denote rejections at the 5 and 10 percent levels respectively.

Table C6

Results of Significance Tests from Logit Equations for the Choice of a Joker Card
Professional Players

Estimating Equation: $J = G[a_0 + a_1 lag(J) + a_2 lag2(J) + b_0 J^* + b_1 lag(J^*) + b_2 lag2(J^*) + c_1 lag(J) lag(J^*) + c_2 lag2(J^*) lag2(J)]$

Null Hypothesis	Player Pairs Whose Behavior Allows Rejection of the Null Hypothesis at:		
		5 percent level	10 percent level
(1). $a_1 = a_2 = b_0 = b_1 = b_2 = c_1 = c_2 = 0$	Row:	2,5,10,14	2,5,10,11,14
	Column:	2,16	2,16
(2). $a_1 = a_2 = 0$	Row:	1,2,5	1,2,3,5,8,14,17
	Column:	2,9,16	2,9,16
(3). $b_1 = b_2 = c_1 = c_2 = 0$	Row:	2,11,17	2,10,11,17
	Column:	2	2,13,18
(4). $c_1 = c_2 = 0$	Row:	2,11	2,11,18
	Column:	2	2,13
(5). $b_1 = b_2 = 0$	Row:	2,11,13	2,11,13
	Column:	-	-
(6). $b_0 = 0$	Row:	-	-
	Column:	-	-

Notes: The symbols J and J* denote the choice of a joker card by a player and by his opponent respectively. The terms "lag" and "lag2" refer to the strategies previously followed in the ordered sequence. The function G[x] denotes the function exp(x)/[1+exp(x)]. Rejections are based on likelihood-ratio tests.

Table D1

Relative Frequencies of Card Choices in O'Neill's Experiment
College Students

A. Frequencies

		Column Player Choice				Marginal Frequencies for Row Player:
		1	**2**	**3**	***J***	
	1	0.045 (0.040) [0.003]	0.042 (0.040) [0.003]	0.040 (0.040) [0.003]	0.079 (0.080) [0.004]	0.205 (0.200) [0.006]
Row Player Choice	**2**	0.044 (0.040) [0.003]	0.046 (0.040) [0.003]	0.038 (0.040) [0.003]	0.080 (0.080) [0.004]	0.207 (0.200) [0.006]
	3	0.042 (0.040) [0.003]	0.034 (0.040) [0.003]	0.046 (0.040) [0.003]	0.075 (0.080) [0.004]	0.196 (0.200) [0.006]
	J	0.076 (0.080) [0.004]	0.084 (0.080) [0.004]	0.078 (0.080) [0.004]	0.154 (0.160) [0.006]	0.392 (0.400) [0.008]
Marginal Frequencies For Column Player:		0.206 (0.200) [0.006]	0.205 (0.200) 0.006	0.202 (0.200) [0.006]	0.387 (0.400) [0.008]	

B. Win Percentages

Observed Row Player Win Percentage:	0.3915
Minimax Row Player Win Percentage:	0.4000
Minimax Row Player Win Std. Deviation:	0.0077

Notes: In Panel A, numbers in parentheses represent Minimax predicted relative frequencies, and numbers in brackets represent standard deviations for observed relative frequencies under the Minimax hypothesis. In Panel B, Minimax Row Player Win Percentage and Std. Deviation are the mean and the std. deviation of the observed row player mean percentage win under the Minimax hypothesis.

Table D2

Relative Frequencies of Card Choices in O'Neill's Experiment by Player Pair
College Students

Pair #	Row Player (R) Choice				Column Player (C) Choice				Rejection of Minimax models	
	1	2	3	*J*	1	2	3	*J*	at 5%:	at 10%:
1	0.225	0.270	0.195	0.310	0.140	0.205	0.260	0.395	a,c,R2,RJ,C1,C3	a,b,c,R2,RJ,C1,C3
2	0.205	0.180	0.160	0.455	0.185	0.205	0.210	0.400		RJ
3	0.200	0.205	0.215	0.380	0.230	0.190	0.225	0.355		
4	0.145	0.215	0.155	0.485	0.175	0.180	0.225	0.420	a,R1,RJ	a,c,R1,R3,RJ
5	0.135	0.190	0.235	0.440	0.195	0.175	0.195	0.435	R1	a,R1
6	0.185	0.215	0.235	0.365	0.215	0.230	0.230	0.325	CJ	CJ
7	0.230	0.185	0.215	0.370	0.200	0.150	0.165	0.485	CJ	b,C2,CJ
8	0.195	0.225	0.165	0.415	0.185	0.225	0.185	0.405		
9	0.150	0.215	0.200	0.435	0.200	0.215	0.195	0.390		R1
10	0.280	0.260	0.200	0.260	0.250	0.185	0.210	0.355	a,c,R1,R2,RJ	a,c,R1,R2,RJ,C1
11	0.195	0.175	0.180	0.450	0.225	0.260	0.205	0.310	b,C2,CJ	b,C2,CJ
12	0.280	0.210	0.180	0.330	0.215	0.220	0.175	0.390	a,R1,RJ	a,c,R1,RJ
13	0.175	0.195	0.195	0.435	0.200	0.200	0.210	0.390		
14	0.170	0.230	0.260	0.340	0.195	0.195	0.205	0.405	R3	a,R3,RJ
15	0.140	0.210	0.200	0.450	0.195	0.205	0.200	0.400	R1	R1
16	0.245	0.195	0.190	0.370	0.225	0.215	0.160	0.400		R1
17	0.195	0.160	0.200	0.445	0.210	0.205	0.200	0.385		
18	0.265	0.210	0.185	0.340	0.215	0.205	0.180	0.400	R1	a,R1,RJ
19	0.300	0.185	0.190	0.325	0.255	0.240	0.195	0.310	a,b,c,R1,RJ,C1,CJ	a,b,c,R1,RJ,C1,CJ
20	0.195	0.205	0.165	0.435	0.215	0.195	0.205	0.385		

Notes: The pairs of capital letters denote rejection of the Minimax binomial model for a given card (1,2,3,J) for a player (R,C).
[a] denotes rejection of Minimax multinomial model for all cards chosen by the row player based on Pearson statistic and $\chi^2(3)$.
[b] denotes rejection of Minimax multinomial model for all cards chosen by the column player based on Pearson statistic and $\chi^2(3)$.
[c] denotes rejection of Minimax multinomial model for all cards chosen by both players based on Pearson statistic and $\chi^2(6)$.

Table D3

Observed and Expected Win Percentages under the Independence Hypothesis

Pair #	Observed Row Win %	Expected Row Win %	Mixture Effect	Correlated Play Effect
1	0.3600	0.4024	0.0024	-0.0424
2	0.3750	0.4006	0.0006	-0.0256
3	0.3750	0.4015	0.0015	-0.0265
4	0.3900	0.4035	0.0034	-0.0135
5	0.4250	0.4024	0.0024	0.0226
6	0.4150	0.4040	0.0040	0.0110
7	0.4000	0.3947	-0.0053	0.0053
8	0.4400	0.3989	-0.0011	0.0411
9	0.3700	0.3991	-0.0009	-0.0291
10	0.3850	0.4095	0.0095	-0.0245
11	0.4000	0.3927	-0.0073	0.0073
12	0.4000	0.3995	-0.0005	0.0005
13	0.4000	0.3994	-0.0007	0.0007
14	0.3900	0.3991	-0.0009	-0.0091
15	0.3500	0.3997	-0.0003	-0.0497
16	0.4000	0.3986	-0.0015	0.0015
17	0.3650	0.3989	-0.0011	-0.0339
18	0.3400	0.3987	-0.0013	-0.0587
19	0.4900	0.4086	0.0085	0.0815
20	0.3600	0.3992	-0.0008	-0.0392
Mean Absolute Value:			0.0027	0.0262
Chi-Square Statistic:			0.2462	18.7426
Chi-Square Probability Value:			0.9999	0.5386

Notes: Expected winning percentages assume independent play with mixtures observed over all 200 games for each pair. Mixture effects are defined as the difference between the values in column 2 and 0.40. Correlation effects are defined as the difference between columns 1 and 2.

Table D4 – Testing that College Students Equate their Strategies' Payoffs to the Equilibrium Rates

		Mixtures		Win Rates			
Pair #	Player	Joker	Non-Joker	Joker	Non-Joker	Pearson	*p*-value
1	R	0.310	0.690	0.371	0.355	8.253	0.041** ‡
	C	0.395	0.605	0.709	0.595	3.885	0.274 †
2	R	0.455	0.545	0.352	0.394	3.542	0.315
	C	0.400	0.600	0.600	0.642	0.868	0.833
3	R	0.380	0.620	0.382	0.371	0.885	0.829
	C	0.355	0.645	0.592	0.643	2.795	0.424
4	R	0.485	0.515	0.351	0.427	7.493	0.058*
	C	0.420	0.580	0.595	0.621	0.542	0.910
5	R	0.440	0.560	0.375	0.464	3.385	0.336
	C	0.435	0.565	0.621	0.540	2.795	0.424
6	R	0.365	0.635	0.438	0.402	1.431	0.698
	C	0.325	0.675	0.508	0.622	6.875	0.076*
7	R	0.370	0.630	0.432	0.381	1.250	0.741
	C	0.485	0.515	0.670	0.534	10.035	0.018**
8	R	0.415	0.585	0.349	0.504	6.274	0.099*
	C	0.405	0.595	0.642	0.504	5.135	0.162
9	R	0.435	0.565	0.425	0.327	3.608	0.307
	C	0.390	0.610	0.526	0.697	6.670	0.083*
10	R	0.260	0.740	0.558	0.324	24.195	0.000** ‡
	C	0.355	0.645	0.592	0.628	2.156	0.541
11	R	0.450	0.550	0.311	0.473	7.639	0.054* †
	C	0.310	0.690	0.548	0.623	7.639	0.054*
12	R	0.330	0.670	0.515	0.343	9.097	0.028* †
	C	0.390	0.610	0.564	0.623	0.764	0.858
13	R	0.435	0.565	0.391	0.407	1.076	0.783
	C	0.390	0.610	0.564	0.623	0.764	0.858
14	R	0.340	0.660	0.324	0.424	4.764	0.190
	C	0.405	0.595	0.728	0.529	8.104	0.044*
15	R	0.450	0.550	0.389	0.318	4.948	0.176
	C	0.400	0.600	0.563	0.708	6.337	0.096*
16	R	0.370	0.630	0.473	0.357	3.281	0.350
	C	0.400	0.600	0.563	0.625	0.781	0.854
17	R	0.445	0.555	0.371	0.360	2.712	0.438
	C	0.385	0.615	0.571	0.675	3.378	0.337
18	R	0.340	0.660	0.368	0.326	6.587	0.086* ‡
	C	0.400	0.600	0.688	0.642	3.420	0.331
19	R	0.325	0.675	0.400	0.533	15.937	0.001** ‡
	C	0.310	0.690	0.581	0.478	16.626	0.001** ‡
20	R	0.435	0.565	0.368	0.354	2.368	0.500
	C	0.385	0.615	0.584	0.675	3.201	0.362

Notes: ** and * denote rejections at the 5 and 10 percent levels respectively. ‡ and † denote the players for whom rejections at the 5 and 10 percent levels of the tests of equality of winning probabilities to the equilibrium rate are also found when the four cards are treated individually.

Table D5

Runs Tests in O'Neill's Experiment
College Students

Pair	Player	Choices: Joker	Choices: Non-Joker	Runs r^j	$F(r^j\text{-}1)$	$F(r^j)$
1	R	62	138	102	0.995**	0.996
	C	79	121	92	0.226	0.271
2	R	91	109	95	0.208	0.251
	C	80	120	92	0.209	0.252
3	R	76	124	92	0.287	0.338
	C	71	129	107	0.985**	0.990
4	R	97	103	99	0.366	0.421
	C	84	116	111	0.961*	0.972
5	R	88	112	103	0.663	0.715
	C	87	113	131	0.999**	1.000
6	R	73	127	99	0.766	0.813
	C	65	135	99	0.942	0.961
7	R	74	126	112	0.996**	0.997
	C	97	103	94	0.146	0.182
8	R	83	117	107	0.889	0.915
	C	81	119	90	0.123	0.155
9	R	87	113	102	0.624	0.677
	C	78	122	96	0.461	0.518
10	R	52	148	82	0.747	0.790
	C	71	129	103	0.937	0.956
11	R	90	110	105	0.740	0.785
	C	62	138	92	0.796	0.834
12	R	66	134	114	0.999**	1.000
	C	78	122	88	0.099	0.126
13	R	87	113	94	0.201	0.244
	C	78	122	89	0.126	0.161
14	R	68	132	72	0.001	0.002**
	C	81	119	84	0.021	0.029*
15	R	90	110	93	0.141	0.176
	C	80	120	112	0.984**	0.989
16	R	74	126	83	0.037	0.052
	C	80	120	101	0.696	0.747
17	R	89	111	112	0.954*	0.966
	C	77	123	109	0.973*	0.981
18	R	68	132	104	0.980**	0.986
	C	80	120	89	0.104	0.135
19	R	65	135	91	0.605	0.673
	C	62	138	91	0.737	0.796
20	R	87	113	107	0.850	0.882
	C	77	123	98	0.606	0.660

Notes: ** and * denote rejections at the 5 and 10 percent levels respectively.

Table D6

Results of Significance Tests from Logit Equations for the Choice of a Joker Card
College Students

Estimating Equation: $J = G[a_0 + a_1 lag(J) + a_2 lag2(J) + b_0 J^* + b_1 lag(J^*) + b_2 lag2(J^*) + c_1 lag(J) lag(J^*) + c_2 lag2(J) lag2(J^*)]$

Null Hypothesis	Player Pairs Whose Behavior Allows Rejection of the Null Hypothesis at		
		5 percent level	10 percent level
(1). $a_1 = a_2 = b_0 = b_1 = b_2 = c_1 = c_2 = 0$	Row:	5,10,12	5,7,9,10,12,14,20
	Column:	5,10,12	4,5,6,10,12,14
(2). $a_1 = a_2 = 0$	Row:	9,12,15	9,10,12,14,15
	Column:	5,14,20	4,5,14,20
(3). $b_1 = b_2 = c_1 = c_2 = 0$	Row:	-	-
	Column:	-	10,12
(4). $c_1 = c_2 = 0$	Row:	-	8,9,15,20
	Column:	-	12
(5). $b_1 = b_2 = 0$	Row:	15	15
	Column:	12	10,12
(6). $b_0 = 0$	Row:	10,12,14	4,10,12,14,16
	Column:	6,10,14	4,6,10,14,16

Notes: The symbols J and J* denote the choice of a joker card by a player and by his opponent respectively. The terms "lag" and "lag2" refer to the strategies previously followed in the ordered sequence. The function G[x] denotes the function exp(x)/[1+exp(x)]. Rejections are based on likelihood-ratio tests.

Table E - Summary Statistics in Penalty Kick's Experiment

		Equilibrium	Professional Soccer Players	College Students with Soccer Experience	College Students with No Soccer Experience
I. Aggregate Data					
Row Player frequencies	*L*	0.363	0.333	0.392	0.401
	R	0.636	0.667	0.608	0.599
Column Player frequencies	*L*	0.454	0.462	0.419	0.397
	R	0.545	0.538	0.581	0.603
Row Player Win percentage		0.7909	0.7947	0.7927	0.7877
(std. deviation)		(0.0074)			
II. Number of Individual Rejections of Minimax Model at 5 (10) percent					
Binomial Model for Marginal Frequencies					
All Players		2 (4)	2 (4)	2 (3)	6 (22)
Multinomial Model for Pair Frequencies					
All Players		1 (2)	0 (2)	2 (2)	6 (9)
III. Mixture and Correlated Effects					
Mean Mixture effect Chi-statistic		---	0.0059	0.0272	0.1139
p-value		---	0.9999	0.9999	0.9999
Mean Correlated effect Chi-statistic		---	3.2343	3.8093	22.3481
p-value		---	0.9999	0.9999	0.3219
IV. Equality of Success Rates Across Strategies and to the Equilibrium Rate					
Rejections at 5 (10) percent		2 (4)	0 (3)	2 (2)	9 (13)
Aggregate Pearson tests					
All Row players statistic		---	40.002	43.294	108.652
p-value		---	0.9781	0.9487	0.000123
All Column players statistic		---	32.486	56.537	113.102
p-value		---	0.9985	0.6030	0.000041
V. Runs tests for 40 players					
Rejections at 5 (10) percent		2 (4)	2 (4)	3 (5)	7 (13)
VI. Logit Equations at the Player Level for 40 players					
Rejections of stationary binomial process at 5 (10) percent		---	2 (4)	4 (7)	12 (14)

Note: The rejections of the stationary binomial process in Panel VI refer to the joint test that all coefficients other than the constant term are equal to zero in the equations specified in section 3.

Table F - Summary Statistics in O'Neill's Experiment

		Equilibrium	Professional Soccer Players	College Students with Soccer Experience	College Students with No Soccer Experience	O'Neill [1]
I. Aggregate Data						
Row Player frequencies	*1*	0.200	0.201	0.203	0.206	0.221
	2	0.200	0.203	0.197	0.206	0.215
	3	0.200	0.198	0.197	0.196	0.203
	J	0.400	0.398	0.403	0.392	0.362
Column Player frequencies	*1*	0.200	0.200	0.199	0.206	0.226
	2	0.200	0.198	0.198	0.205	0.179
	3	0.200	0.204	0.203	0.201	0.169
	J	0.400	0.398	0.400	0.387	0.426
Row Player Win percentage		0.400	0.394	0.403	0.391	0.410
(std. deviation)		(0.007)				
II. Number of Individual Rejections of Minimax Model at 5 (10) percent						
Row Player (All Cards)		1 (2)	1 (1)	2 (2)	5 (8)	6 (na)
Column Player (All Cards)		1 (2)	1 (1)	2 (2)	2 (4)	9 (na)
Both Players (All Cards)		1 (2)	2 (2)	1 (2)	3 (5)	9 (na)
All Cards		8 (16)	8 (18)	11 (18)	23 (31)	35 (na)
III. Mixture and Correlated Effects						
Mean Mixture effect *p*-value		---	0.999	0.999	0.999	0.999
Mean Correlated effect *p*-value		---	0.996	0.999	0.538	0.026
IV. Equality of Success Rates Across Strategies and to the Equilibrium Rate using *NJ* and *J* [2]						
A. Using equilibrium frequencies and success probabilities (3 degrees of freedom at individual level)						
Rejections at 5 (10) percent		2 (4)	3 (6)	3 (6)	5 (15)	22 (25)
Aggregate Pearson tests						
All Row players *p*-value		---	0.715	0.514	0.000009	$6.78 \cdot 10^{-17}$
All Column players *p*-value		---	0.654	0.959	0.0042	$1.90 \cdot 10^{-21}$
B. Using maximum likelihood estimates (1 degree of freedom at individual level)						
Rejections at 5 (10) percent		2 (4)	2 (4)	3 (6)	8 (10)	10 (15)
Aggregate Pearson tests						
All Row players *p*-value		---	0.404	0.221	0.005	$4.93 \cdot 10^{-8}$
All Column players *p*-value		---	0.298	0.387	0.002	$1.45 \cdot 10^{-8}$
V. Runs Tests						
Rejections at 5 (10) percent		2 (4)	2 (4)	3 (5)	7 (12)	15 (19)
VI. Logit Equations for Stationary Binomial Process [3]						
Rejections at 5 (10) percent		---	6 (7)	4 (9)	6 (13)	31 (na)

1. The results for O'Neill come from Brown and Rosenthal (1990) and Walker and Wooders (2001), where "na" means that the corresponding estimate was not reported by the authors and may not be computed from the data they report. O'Neill's (1987) experiment involves 25 pairs, rather than 20 pairs, and 105 repetitions instead of 200. Hence, the number of expected rejections under Minimax at a given percentage level in the original O'Neill's experiment is 1.25 greater than those reported in the first column, and the std. deviation for observed relative frequencies under Minimax play in Panel I is 0.009, rather than 0.007.
2. In O'Neill's original experiment there are two pairs that represent extreme outliers. When these are ignored, the *p*-values in panel A remain very low ($1.2 \cdot 10^{-9}$ and $1.7 \cdot 10^{-12}$, respectively).
3. Rejections refer to the joint test that all coefficients except the constant term are zero in the equations specified in section 3.