EXPERIENTIA DOCET: PROFESSIONALS PLAY MINIMAX IN LABORATORY EXPERIMENTS

BY IGNACIO PALACIOS-HUERTA AND OSCAR VOLIJ[1]

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 2×2 game that is formally identical to a strategic interaction situation that they face in their natural environment. Consistent with their behavior in the field, they play very close to the equilibrium of the game. In particular, (i) they equate their strategies' payoffs to the equilibrium ones and (ii) they generate sequences of choices that are serially independent. In sharp contrast, however, we 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 4×4 zero-sum game, a game that none of the subjects has encountered previously, and find the same differences in the behavior of these two pools of subjects. The transfer of skills and experience from the familiar field to the unfamiliar laboratory observed for professional players is relevant 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.

KEYWORDS: Laboratory experiments, minimax, experience, cognition.

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. (Hume (1739))

1. INTRODUCTION

AN IMPORTANT ISSUE in economic research that relies on data collected in a laboratory is how applicable are the insights gained for predicting behavior in natural environments. This paper addresses this issue for situations that involve strategic interaction. Game theory is, in fact, one of the areas where experimental data from the laboratory are often used to inform theoretical developments.[2] One reason for this is that nature does not always create the

---

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

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. This is important because game-theoretic predictions are often sensitive to changes in these variables. However, as Harrison and List (2004, pp. 1009–1011) point out, "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 concerns about the extent to which laboratory results may provide insights into field behavior demand more elaborate experiments.[3]

In this paper we conduct a conventional experiment in which a nonstandard pool of subjects plays a game whose unique equilibrium involves mixed strategies. Our idea is to use professional soccer players to develop an "artefactual field experiment" to study an aspect of games with mixed-strategy equilibria that was not studied previously.[4] Soccer has three unique features which make it especially suitable for this purpose: (i) Professional soccer players face a *simple* strategic interaction that is governed by very detailed rules: a penalty kick. (ii) The formal structure of this interaction 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 the field can be transferred to the laboratory. Further, the extent to which field and laboratory behavior differ can indicate whether laboratory findings are reliable for predicting field behavior.

We proceed as follows. We first analyze the behavior of professional soccer players in a laboratory setting playing a simultaneous two-person zero-sum 2×2 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. To test our methodological hypothesis, we also implement exactly the same controlled laboratory experiment with subjects drawn from a standard subject pool of college students with no soccer experience.

---

[3]See Harrison (2005), 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.

[4]We use the term suggested in Harrison and List's (2004) classification of experiments.

Palacios-Huerta (2003) found that the behavior of professional players in the soccer field was consistent with equilibrium play: (i) their winning probabilities were statistically identical across strategies and (ii) their choices were serially independent.[5] The results we obtain in this paper can be summarized as follows. We find that professional players' behavior continues to be rather consistent with the implications of equilibrium in the entirely different setting of a lab. 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: their distribution of play is significantly different from the equilibrium one and they generate sequences that exhibit negative autocorrelation. We interpret these results as evidence that professionals transfer their skills across these vastly different environments and circumstances. 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 has consistently found that subjects are unable to generate independent and identically distributed (i.i.d.) sequences in the laboratory, but rather exhibit a significant bias against repeating the same choice.[6] We find, however, that professional soccer players appear to generate random sequences in the lab while college students do not.

To evaluate whether professional players behave differently in a zero-sum game they have not encountered previously in any setting, we asked them to play the 4×4 two-person simultaneous game developed by O'Neill (1987) and further studied by Brown and Rosenthal (1990), Shachat (2002), and Walker and Wooders (2001). We also compare their behavior with that of college students. Consistent with this literature, the results show that students behave in a manner that is far from 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 they generate sequences of choices that are not random. In sharp contrast to this behavior, we find that professional soccer players play quite close to equilibrium in most dimensions. Interestingly, one exception we find partially comes from a component of this game that lacks a familiar context.

We interpret the results that professionals whose play in the field is consistent with equilibrium also behave close to equilibrium in a laboratory setting as supporting the idea that the vast differences in environments do not undermine the skills these subjects use in the field. This interpretation is strengthened by

---

[5]See also Chiappori, Levitt, and Groseclose (2002) for further evidence in support of equilibrium behavior.

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

the fact that payoffs in the lab are lower than in real life and that one of the games played in the lab is entirely unfamiliar to the subjects. The evidence, therefore, suggests that the game-theoretic equilibrium predictions may have greater empirical content than previously thought for explaining behavior in *both* natural and experimental settings. Further, the fact that the behavior of professional soccer players is distinctly different from that of college students, the subject pool typically studied in a large experimental literature, indicates that the nature of the subject pool may be a critical ingredient of the laboratory experiment for predicting field behavior.

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. While perfectly competitive games do not represent the entire universe of strategic games involving mixed strategies, they are considered a "vital cornerstone" of game theory (e.g., Aumann (1987), Binmore, Swierzbinski, and Proulx (2001)). Indeed, zero-sum games can be regarded as the branch of game theory with the most solid theoretical foundations.[7] From this perspective, the positive results we find lend support to a fundamental result of game theory in a setting where the small number of existing results have been mainly negative. Last, from the viewpoint of the literature on cognition and similarity as a basis for inductive reasoning,[8] the results support the hypothesis that cognitive skills may exist beyond those that subjects are aware of in the context of games involving mixed strategies, and that they can be transferred to the highly unfamiliar environment of the lab.

2. EXPERIMENTAL PROCEDURES

We implement two different zero-sum games, each with two subject pools: professional soccer players and college students. The experiments were conducted during the period November 2003–October 2004 at the Universidad del País Vasco in Bilbao (Spain). Each of the two zero-sum games was played by a different set of 40 professional soccer players working in twenty pairs and 40 college students with no soccer experience working in twenty pairs. We also recruited two additional sets of 40 college students with soccer experience at the amateur level, one for each of the two games, for one of the extensions of the analysis that will be discussed later. In what follows, we explain the recruiting process for these 240 subjects and other aspects of the experimental procedure, and then describe the experimental designs of the two games.

---

[7]Within the class of zero-sum games even the less stringent concept of correlated equilibrium coincides with the set of minimax strategies. In this sense, the theory of minimax is one of the less controversial ones from a theoretical point of view.

[8]See, for instance, Hume (1739), Gilboa and Schmeidler (2001), Gigerenzer and Todd (1999), Selten (1998), Simon (1983), and other references therein.

2.1. *Subjects*

Each subject participated in only one type of game and one session. Sessions lasted about an hour, and subjects received their winnings as payment.

PROFESSIONAL PLAYERS: These subjects were recruited from professional soccer clubs in Spain. As in many other 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 were taken from clubs in the north of Spain, a region with a high density of professional teams.

Eighty male soccer players (40 kickers and 40 goalkeepers) were recruited from these teams by telephone and by interviews during their daily practice sessions. 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.[10] Forty kicker–goalkeeper pairs were formed randomly using the last two digits of their national ID card, the only requirement being that subjects who 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 from the Universidad del País Vasco in Bilbao by making visits to different undergraduate classes. Subjects majoring in economics or mathematics were excluded from the sample. Half of the subjects had no soccer experience. The other half had soccer experience at the amateur level since it was required that they should be currently participating in regular league competitions in regional amateur divisions (the Tercera Division and below).[11] These leagues adhere exactly to the same structure and calendar schedule, and are governed by the same rules (FIFA (2005)) as professional leagues.

---

[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. The Tercera Division comprises 240 teams. 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 the Tercera Division and these regional leagues.

[10]The average age in the sample is 26.5 years, and the average number of years of education is 11.2. No player who had played professionally for less than 2 years at the time of the experiment was recruited. Data on wages and salaries on individual players are not publicly available, but estimates from Marca (2005) and Deloitte and Touche (2005) indicate that wage expenditures represent between 60 and 75 percent of revenue for most clubs. This means that for a typical squad of 25 players, the average yearly wage is about 2 million dollars in the Primera Division and 0.5 million dollars in the Segunda Division A. These amounts exclude other sources of revenues such as endorsements.

[11]The average age in the sample is 20.7 years, and the average number of years of education is 15.1. There are no statistical differences between the two pools of college subjects.

Pairs were formed randomly using the last two digits of their national ID card. For subjects with soccer experience, those who 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*

In soccer, a penalty kick is awarded against a team that commits a punishable offense inside its own penalty area while the ball is in play. The *Official Laws of the Game* (FIFA (2005)) describe in detail the simple rules that govern this strategic interaction. Each penalty kick involves two players: a kicker and a goalkeeper. In a 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 the goalkeeper's movement time to any possible path of the ball. 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 the players choose their strategies.

The clarity and simplicity of these rules suggest not only that the penalty kick 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:

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

where π_{ij} denotes the kicker's probability 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:[12]

1. The probability that a goal will be scored must be the same across each player's strategies and be equal to the equilibrium scoring probability, namely $p = (\pi_{LR}\pi_{RL} - \pi_{LL}\pi_{RR})/(\pi_{LR} - \pi_{LL} + \pi_{RL} - \pi_{RR})$.

2. Each player's choices must be serially independent. Hence, a player's choices must be independently drawn from a random process and should not depend on his own previous play, on the opponent's previous play, or on any other previous actions and outcomes.

---

[12]When a zero-sum two-outcome game is repeated a finite number of times, the only equilibrium of the resulting supergame consists of the repetition of the equilibrium of the one-shot game at every round independently of players' risk preferences (see Wooders and Shachat (2001)).

Using data on over a thousand penalty kicks during a 5-year period in three countries, Palacios-Huerta (2003) found strong support for the two implications of this 2×2 model. We adopt this model and bring it to the laboratory. The payoffs we use in the experiment are

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

They are derived from a sample of 2,717 penalty kicks collected from European professional leagues during the period 1995–2004.[13] No other field references and no soccer terminology that might trigger any type of psychological reaction were used in the experiment.[14] In particular, subjects were 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 follow as closely as possible those of O'Neill (1987). The players sit opposite each other at a table. Kickers play the role of row player and goalkeepers play the role of column player. Each player holds two cards (A and B) with identical backs. A large board across the table prevents the players from seeing the backs of their opponents' cards. The experimenter hands out one page with the following instructions (in Spanish), which he then reads 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 tens 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 00 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.

[13]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).

[14]The choice of parameters can add, to some extent, a 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 a laboratory setting.

● 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:

	A	B
A	.60	.95
B	.90	.70

Are there any questions?"

Thus, the game was presented with the help of a matrix, and the subjects learned the rules after a few rounds of practice. The unique mixed-strategy equilibrium of this game dictates that row and column players choose the *A* card with probabilities 0.3636 and 0.4545, respectively. The subjects 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 would play. On the few occasions when they made an error announcing the winner, the experimenter corrected them.

A typical session lasted about 1 hour and 15 minutes, proceeding at about two hands per minute. From the perspective of the response times study of Rubinstein (2007) on instinctive and cognitive reasoning, it is of interest to note that professionals took on average 70 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 sit opposite each other at a table. Each player holds four cards with identical backs. A large board across the table prevents the players from seeing the backs of their opponents' cards. The experimenter hands out one page with the following instructions (in Spanish) to the participants, which he then reads 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)."

The original instructions in Spanish for both this experiment and the penalty kick experiment are available in the online supplement (Palacios-Huerta and Volij (2008)). Thus, in this case the game was presented without the help of a matrix and subjects learned the rules by practice. The payoff structure of the game is

	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 chose the red, brown, purple, and green cards with probabilities 0.2, 0.2, 0.2, and 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 would play. If they happened to make an error in determining the winner, the experimenter corrected them.

A typical session lasted slightly more than 1 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.

There are several differences between our design and that of O'Neill. For one thing, our subjects engage in 200 stage games instead of 105. Second, we rename the elements of the action space to be {*Red*, *Brown*, *Purple*, *Green*}, as in Shachat (2002), rather than using {*Ace*, *Two*, *Three*, *Joker*}. This is done to avoid the previously observed *Ace* bias.[15] Nonetheless, to avoid confusion and to facilitate comparison with the literature, actions will be referred to by the names used in O'Neill's experiment for the remaining exposition of the paper, namely 1 (*Ace*) for *Red*, 2 (*Two*) for *Brown*, 3 (*Three*) for *Purple*, and J (*Joker*) for *Green*. A last difference is that we use much greater stage game payoffs (the winner receives 1 euro for a win rather than 5 cents; that is about 1.30 dollars using the exchange rate at the time the experiment took place) and we do not provide any initial endowments to the players.

3. EXPERIMENTAL 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.

---

[15]See O'Neill (1987) and Brown and Rosenthal (1990).

3.1. *Penalty Kick Experiment*

3.1.1. *Professional players*

Table I presents aggregate statistics describing the outcomes of the experiment. 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. The minimax relative frequencies appear in parentheses and the standard deviation for the observed relative frequencies under the minimax hypothesis appears in brackets. 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.

The pattern of observed relative frequencies for each pair of choices shows a general consistency with the minimax model in that they all are within 1 to 2 percentage points of the predicted frequencies. Likewise, the marginal frequencies of actions for the players are extremely close to the minimax predic-

TABLE I

RELATIVE FREQUENCIES OF CHOICES AND WIN PERCENTAGES IN PENALTY KICK EXPERIMENT
WITH PROFESSIONAL PLAYERS[a]

A. Frequencies		Column Player Choice		Marginal frequencies for row player
		L	R	
	L	0.152 (0.165) [0.0068]	0.182 (0.198) [0.0073]	0.333 (0.364) [0.0088]
Row Player Choice	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

[a]In panel A the numbers in parentheses represent minimax predicted relative frequencies, whereas those 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 standard deviation of the observed row player mean percentage win under the minimax hypothesis.

tions for the column player. Row players, on the other hand, choose frequencies 0.333 for L and 0.667 for R, which, while close to the minimax predictions, are statistically different from them.[16] The observed aggregate row player win frequency (0.7947) is less than 1 standard deviation away from the theoretically expected value (0.7909). Although the aggregate mixture of the row players is statistically different from the equilibrium one, the difference is minuscule. Indeed, row players chose L with probability 0.33, while the equilibrium prescribes 0.36. Also, if column players played the best response to row players' actual mix, their success rate would increase from 20.9 percent to only 21.6 percent.

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 II reports the relative frequencies of choices for each of the twenty pairs in the sample and some initial tests of the model.

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 row and column players, respectively. We should then expect a binomial test of conformity with minimax play to reject the null hypothesis for two players at the 5 percent significance level, and four players at the 10 percent level. The results show that indeed these are precisely the number of rejections at those confidence levels.

These initial findings support the hypothesis that professional soccer players play very close to the equilibrium of the game, though not perfectly. However, since equilibrium behavior also implies that action combinations should be realizations of *independent* drawings of a multinomial distribution, further support is needed. To test whether the players' actions are correlated, we perform the following test. 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 II 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-squared 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 minimax play we would expect to reject the null hypothesis for one and two pairs at the 5 and 10 percent significance levels. We find 0 and 2 rejections, respectively.

Summing up, even though the observed aggregate frequency for the row players is statistically different from the equilibrium predictions, this initial evidence lends substantial support to the minimax hypothesis. Our next task is to test more closely the implications of the equilibrium of the game.

---

[16]Indeed, the p-value of the null hypothesis that row players choose the equilibrium frequencies is 0.06 percent, for column players it is 41 percent, and for both players it is 0.48 percent.

TABLE II

MARGINAL FREQUENCIES AND ACTION PAIR FREQUENCIES IN PENALTY KICK EXPERIMENT
WITH PROFESSIONAL PLAYERS[a]

	Marginal Frequencies		Pair Frequencies				
Pair #	Row L	Column L	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

[a]The ** 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 frequencies.

IMPLICATION 1—Winning Rates and the Distribution of Play: Minimax play implies that the success probabilities of each action will be the same for each player and be 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 action–outcome frequencies associated with the equilibrium. Table III reports the relative frequencies of action–outcome combinations observed for each of the row and column players in the sample. Using the absolute frequencies that correspond to these entries, we can then implement a chi-squared test of conformity with minimax play. This test would be identical to the one performed in Table II 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.

The results of the test show that the null hypothesis is rejected for no player at the 5 percent significance level and for three players at the 10 percent significance level, both cases being fewer than the expected number of rejections, 2 and 4, respectively. Hence, at the individual level the hypothesis that scoring

TABLE III

TESTING THAT PROFESSIONAL PLAYERS EQUATE THEIR STRATEGY PAYOFFS TO THE
EQUILIBRIUM RATES[a]

Pair #	Player	L		R		Pearson Statistic	p-Value
		Success	Fail	Success	Fail		
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

[a]The ** and * denote rejections at the 5 and 10 percent levels, respectively.

probabilities are identical both across strategies and to the equilibrium rate cannot be rejected for most players at conventional significance levels.

The question of whether behavior at the aggregate level is generated from equilibrium play 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 player. Under the null hypothesis, it is distributed as a χ^2 with 60 degrees of freedom for both the set of row players and the set of column players. We find that the Pearson statistics are 40.002 and 32.486, with an associated p-value above 90 percent in both cases.[17] Hence, the null hypothesis that the data for all players are 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 ones.

IMPLICATION 2—The Serial Independence Hypothesis: Another 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. To our knowledge 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 Kahneman (1971), Camerer (1995)). The only possible exception that we are aware of is owing to Neuringer (1986), who explicitly taught subjects to choose randomly after 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.[19] In our case, however, subjects have accumulated their experience in the entirely different environment of a soccer field. Moreover, professional soccer players rarely take penalty kicks in the field in rapid succession, as they are asked to do in the experiment. Instead, there is often a substantial time delay, typically weeks, between subsequent penalties. Whether their skills and experience in the field are useful to generate random sequences in a laboratory setting where stage games are repeated in rapid succession is the question we turn to next.

---

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

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

[19]See Neuringer (2002) for a thorough review of the literature.

To address this question, we consider the "runs test" of serial independence previously performed by Walker and Wooders (2001), which proceeds as follows. Take the sequence of actions chosen by player i in the order in which they occurred $s^i = \{s_1^i, s_2^i, \ldots, s_{ni}^i\}$, 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. A run is defined as a succession of one or more identical actions which are preceded and followed by a different action or no action at all. When the choices s_x^i are serially independent, all the combinations of n_R^i right choices and n_L^i choices out of $n_L^i + n_R^i$ choices are equally probable. In that case, the probability of observing r runs in a sequence of $n_L^i + n_R^i$ action choices, n_L^i and n_R^i right, is known (see Gibbons and Chakraborti (1992)) and given by

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

$$
\text{for } r = 2, 3, \ldots, n_L^i + n_R^i.
$$

Let r^i be the observed number of runs in the sequence s^i. Then the null hypothesis of serial independence will be rejected at the 5 percent confidence level if the probability of r^i or fewer runs is less than 0.025 or if the probability of r^i or more runs is less than 0.025; that is, if $F(r|n_L^i, n_R^i) < 0.025$ or if $1 - F(r - 1|n_L^i, n_R^i) < 0.025$, where $F(r|n_L^i, n_R^i) = \sum_{k=1}^{r} f(k|n_L^i, n_R^i)$ denotes the probability of obtaining r or fewer runs. The results of these tests are shown in Table IV.

We find that the null hypothesis of serial independence is rejected for two players at the 5 percent significance level and for four players at the 10 percent level, precisely the expected number of rejections in both cases under the null hypothesis. These results indicate that, according to this test, the hypothesis that professional soccer players generate random sequences cannot be rejected. They switch strategies neither too often nor too little. Moreover, the number of rejections is remarkably consistent with the theory. This behavior is in sharp contrast to the overwhelming experimental evidence from the psychological and experimental literatures mentioned earlier. It indicates that field skills and years of experience may be 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 found in the laboratory.

TABLE IV
RUNS TESTS IN PENALTY KICK EXPERIMENT WITH PROFESSIONAL PLAYERS[a]

Pair	Player	Choices		Runs		
		R	L	r^i	$F(r^i - 1)$	$F(r^i)$
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

[a]The ** and * denote rejections at the 5 and 10 percent levels, respectively.

This evidence represents the first time that subjects have been found to display statistically significant serial independence in a strategic game in a laboratory setting. Furthermore, together 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 satisfy these two equilibrium conditions in the laboratory in games where players are predicted to choose probabilistic mixtures.

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 V presents aggregate statistics that describe the aggregate outcomes of the experiment.

The aggregate data for these players also seem to conform well to the equilibrium predictions since the observed frequencies appear to be broadly consistent with 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 less than 1 standard deviation away from the expected value. However, a closer look quickly reveals that the observed

TABLE V

RELATIVE FREQUENCIES OF CHOICES AND WIN PERCENTAGES IN PENALTY KICK EXPERIMENT
WITH COLLEGE STUDENTS[a]

A. Frequencies				Marginal frequencies for row player
		Column Player Choice		
		L	R	
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

[a]In panel A the numbers in parentheses represent minimax predicted relative frequencies, whereas those 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 standard deviation of the observed row player mean percentage win under the minimax hypothesis.

TABLE VI

MARGINAL FREQUENCIES AND ACTION PAIR FREQUENCIES IN PENALTY KICK EXPERIMENT
WITH COLLEGE STUDENTS[a]

| | 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**

[a]The ** 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 frequencies.

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.[20] Furthermore, *both* players choose very similar frequencies, roughly 0.40 for L and 0.60 for R. This suggests that these subjects may not appreciate the slight differences in payoffs in the off-diagonal elements of the payoff matrix, differences that in equilibrium induce players to adopt different strategies from the opponent.

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

First, the binomial test for conformity with minimax play indicates that the model is rejected for 6 and 22 players at the 5 and 10 percent levels, respectively. This excessively high amount of rejections, three and more than five

---

[20]The aggregate chi-squared test for the conformity with minimax play based on Pearson goodness of fit has a p-value of 2×10^{-13}.

times greater than those predicted by the equilibrium of the game at those levels, indicates that there will be substantial deviations from equilibrium play in the subsequent tests of the minimax hypothesis. Indeed, using the absolute frequencies that correspond to the observed joint choices reported in the table and their associated minimax probabilities, a chi-squared test for conformity with minimax play indicates that the model is rejected for 6 and 9 pairs at the 5 and 10 percent levels of significance. Under the null hypothesis we would expect only 1 and 2 rejections, respectively.[21]

To sum up, the excessively high amount of rejections in these tests clearly indicates substantial deviations from equilibrium play. Next, we test the minimax predictions more closely.

IMPLICATION 1—Winning Rates and the Distribution of Play: Table VII tests whether the observed distribution of play is equal to the equilibrium distribution using the success rates of each action for each player.

Using the absolute frequencies that correspond to each action–outcome combination, a chi-squared test shows that the minimax multinomial model is rejected for 9 players at the 5 percent significance level and 13 players at the 10 percent level, when the expected number of rejections under the hypothesis of minimax play is 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.

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 it is 113.102, with associated p-values of 1.2×10^{-4} and 4.1×10^{-5}, respectively. Hence, the null hypothesis that the data for all players are generated by equilibrium play is strongly rejected at conventional significance levels.

These results, therefore, indicate that observed behavior is far from the equilibrium one and highly different from professional soccer players' behavior.

IMPLICATION 2—The Serial Independence Hypothesis: To test whether subjects randomize across actions using the same probability distribution at each stage, we implement the runs test of serial independence. The results are given in Table VIII.

The null hypothesis of serial independence is rejected for 7 and 13 players at the 5 and 10 percent significance levels, in each case more than three times

---

[21]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.

TABLE VII

TESTING THAT COLLEGE STUDENTS EQUATE THEIR STRATEGY PAYOFFS TO
THE EQUILIBRIUM RATES[a]

| Pair # | Player | L | | R | | Pearson | |
		Success	Fail	Success	Fail	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**

[a] The ** and * denote rejections at the 5 and 10 percent levels, respectively.

TABLE VIII

RUNS TESTS IN PENALTY KICK EXPERIMENT WITH COLLEGE STUDENTS[a]

Pair	Player	Choices		Runs		
		R	L	r^i	$F(r^i - 1)$	$F(r^i)$
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

[a]The ** and * denote rejections at the 5 and 10 percent levels, respectively.

the number of expected rejections. These findings indicate that college subjects do not 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. Also consistent with past evidence is the fact that in most cases the reason for the rejections is an excessive number of alternations.

Consequently, the results of the tests of serial independence decisively indicate that individuals display statistically significant serial dependence. Together with the results from the tests of equality of winning probabilities, we can conclude that the minimax model is not supported for college students.

3.2. *O'Neill's Experiment*

The differences between professional soccer players and college students are substantial in the penalty kick experiment. Professionals' behavior is very close to the equilibrium of the game, while college students' behavior is far from it. In this section we use a different zero-sum game, namely that introduced in O'Neill (1987), in an attempt to study whether the experience and skills that professional players use in the field are valuable 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 IX presents aggregate statistics that describe observed relative frequencies for each pair of moves and each card. Minimax relative frequencies appear in parentheses and their standard deviations under the minimax hypothesis appear in brackets. The bottom panel reports the observed win frequencies for the row player.

These aggregate data 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 action pairs involving non-*J* cards are in the neighborhood of 0.04, while relative frequencies for pairs involving one *J* card and for the pair involving the two *J* cards are in the neighborhood of 0.08 and 0.16, respectively. The aggregate row player win frequency (0.3945) is less than 1 standard deviation away from the expected value (0.40). Also, a chi-squared 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 90 percent. In addition, the marginal frequencies of actions for the row and column players are extremely close to the minimax predictions. In every case, they are less than 1 standard deviation away.

Table X reports the observed marginal frequencies for each player. Under the minimax hypothesis, each player's chosen actions are a realization of a multinomial distribution with probabilities 0.4 for the *J* card, and 0.2 for each

TABLE IX

RELATIVE FREQUENCIES OF CARD CHOICES AND WIN PERCENTAGES IN O'NEILL'S
EXPERIMENT WITH PROFESSIONAL PLAYERS[a]

A. Frequencies

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

[a] In Panel A the numbers in parentheses represent minimax predicted relative frequencies, whereas those 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 standard deviation of the observed row player mean percentage win under the minimax hypothesis.

of the other cards. The table reports the p-values of the corresponding chi-squared tests with 3 degrees of freedom. The minimax hypothesis also implies that observed action combinations for each pair (not reported here) are a realization of a multinomial distribution that results from the product of the marginals mentioned above. The last column of Table X reports the p-values of the corresponding chi-squared tests with 15 degrees of freedom.

The expected number of rejections in the case of individual players is 2 and 4 at the 5 and 10 percent significance levels. We find that the actual number of players for whom the null hypothesis is rejected is 2 and 3 at these levels. Likewise, at the pair level we would expect 1 and 2 rejections, respectively. We find, however, 2 and 6 rejections. These differences between the individual

TABLE X

RELATIVE FREQUENCIES OF CARD CHOICES IN O'NEILL'S EXPERIMENT WITH PROFESSIONAL PLAYERS AND MINIMAX MULTINOMIAL TESTS[a]

| | Row Player (R) Choices | | | | Column Player (C) Choices | | | | p-Values for Tests of Minimax Multinomial Models | | |
| | | | | | | | | | Row | Column | Both |
Pair #	1	2	3	J	1	2	3	J	Player	Player	Players
1	0.190	0.225	0.290**	0.295**	0.195	0.185	0.210	0.410	0.002‡	0.940	0.018‡
2	0.205	0.215	0.245*	0.335**	0.200	0.205	0.250*	0.345*	0.223	0.257	0.001‡
3	0.210	0.195	0.200	0.395	0.195	0.175	0.205	0.425	0.987	0.804	0.802
4	0.215	0.205	0.180	0.400	0.145**	0.185	0.225	0.445	0.885	0.180	0.126
5	0.180	0.195	0.205	0.420	0.200	0.195	0.210	0.395	0.885	0.987	0.959
6	0.210	0.205	0.185	0.400	0.205	0.185	0.205	0.405	0.950	0.962	0.873
7	0.215	0.215	0.130**	0.440	0.205	0.190	0.205	0.400	0.105†	0.985	0.067†
8	0.195	0.215	0.195	0.395	0.225	0.150*	0.205	0.420	0.962	0.341	0.346
9	0.185	0.195	0.215	0.405	0.205	0.180	0.205	0.410	0.922	0.919	0.844
10	0.175	0.180	0.170	0.475**	0.195	0.195	0.215	0.395	0.192	0.962	0.787
11	0.205	0.190	0.170	0.435	0.250*	0.200	0.205	0.345*	0.651	0.257	0.864
12	0.200	0.200	0.195	0.405	0.195	0.200	0.205	0.400	0.998	0.997	0.986
13	0.215	0.185	0.195	0.405	0.195	0.215	0.190	0.400	0.922	0.950	0.893
14	0.185	0.185	0.205	0.425	0.205	0.290**	0.195	0.310**	0.852	0.007‡	0.062†
15	0.215	0.200	0.170	0.415	0.210	0.185	0.200	0.405	0.744	0.953	0.923
16	0.205	0.195	0.195	0.405	0.195	0.165	0.175	0.465*	0.993	0.263	0.231
17	0.205	0.230	0.190	0.375	0.225	0.215	0.205	0.355	0.720	0.596	0.081†
18	0.210	0.195	0.180	0.415	0.205	0.245*	0.210	0.340*	0.888	0.267	0.080†
19	0.205	0.220	0.235	0.340*	0.170	0.205	0.175	0.450	0.327	0.423	0.192
20	0.195	0.210	0.200	0.395	0.185	0.205	0.180	0.430	0.987	0.777	0.436

[a]The ** and * denote rejection of the minimax binomial model for a given card and player at the 5 and 10 percent levels, respectively. Similarly, ‡ and † denote rejection at those levels of the minimax multinomial model based on Pearson statistic and a $\chi^2(3)$ for all cards chosen by a given row or column player, and based on Pearson statistic and a $\chi^2(15)$ for all cards chosen by both players.

and pair level number of rejections indicate that there is a contemporaneous correlation in some players' choices.

An interesting aspect of these tests is concerned with the distribution of p-values.[22] If players' choices were drawn from the equilibrium mixture, the resulting p-values should be realizations of a uniform distribution $U[0, 1]$ both in the individual and pair level tests. But a look at the distribution of p-values in the individual level tests readily reveals that this is not the case in that the distribution is skewed toward 1. There are far too many values above 0.5 and too few below 0.5.[23] For instance, there are about 14 values too many above 0.85: we would expect 6 such values and we find 20. Thus, while about two-thirds of the values seem consistent with a $U[0, 1]$ distribution, the marginal distributions of one-third of the sample are too close to the equilibrium ones, suggesting that these players do not behave as perfect i.i.d. randomizers. Section 4 discusses in more detail this feature of players' behavior.

On the other hand, the evidence at the pair level is more consistent with strict randomization. If we examine the hypothesis that the joint behavior is a realization of the product of the equilibrium mixtures, the Kolmogorov–Smirnov test of equality of the empirical distribution of p-values and the uniform distribution $U[0, 1]$ results in a statistic equal to 1.06 (p-value $= 0.2105$), which is not large enough to reject the null hypothesis at the 20 percent level.

Based on the above findings, we can say that minimax theory predicts well the individual and joint behavior of our subjects.

IMPLICATION 1—Winning Rates and the Distribution of Play: Table XI 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 3 degrees of freedom given that the game being played is known. The table also indicates the rejections that are obtained when the test is implemented for the individual choices of cards (i.e., when 1, 2, 3, and *J* are treated on an individual basis).

The results show that for the choice of *Joker* and *non-Joker* the null hypothesis is rejected for three players at the 5 percent significance level and six players at 10 percent level, whereas the number of rejections when the test is implemented for the individual card choices is 3 and 4 at these levels, respectively. In both cases, therefore, the number of rejections is very close to the expected number (2 and 4, respectively) according to the null hypothesis.

We also test whether behavior at the aggregate level is generated from equilibrium play. Since the Pearson statistic for the joint test for all row players

---

[22]We are indebted to a referee for pointing this out.

[23]Indeed, if we perform a Kolmogorov–Smirnov test of equality of the empirical distribution of the forty players' p-values to the uniform distribution, we get a statistic equal to 2.43, which is high enough to reject the null hypothesis at virtually any significance level.

TABLE XI

TESTING THAT PROFESSIONAL PLAYERS EQUATE THEIR STRATEGY PAYOFFS TO THE
EQUILIBRIUM RATES IN O'NEILL'S EXPERIMENT[a]

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

[a]The ** and * denote rejections at the 5 and 10 percent levels, respectively; ‡ and † denote rejections at the same levels when the four cards (1, 2, 3, J) are treated on an individual basis.

is 53.351, with an associated *p*-value of 0.715, and for all column players is 55.122, with an associated *p*-value of 0.654, the null hypothesis that the data for all players were generated by equilibrium play cannot be rejected at conventional significance levels.

IMPLICATION 2—The Serial Independence Hypothesis: In Table XII we implement the runs tests to study whether players' choices are serially independent for the choice of *Joker* and *non-Joker* cards. We find that the null hypothesis of serial independence is rejected for two and four players at the 5 and 10 percent significance levels, which, according to the theory, is precisely the number of rejections that we should expect at these levels.

These findings, therefore, support the hypothesis that professional soccer players are able to generate random sequences in the laboratory. Since professionals behave relatively close to the equilibrium behavior of the game according to the previous tests as well, the results are consistent with the idea that field skills and experience can be transferred to this zero-sum game as well.

3.2.2. *College students*

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

The results are presented in Tables XIII–XVI to parallel the presentation of the empirical evidence for the professional soccer players. They can be summarized as follows. Our findings are consistent with those obtained by Brown and Rosenthal (1990), Walker and Wooders (2001), and Shachat (2002) for O'Neill's experiment. Even though aggregate frequency data do 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 1 standard deviation away from the predicted value (Table XIII). The hypotheses that players mix according to the equilibrium distributions are rejected, both at the individual and at the pair level, in an excessively large number of cases (Table XIV). 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 that each observation is generated from equilibrium play is 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 XV). There is also strong evidence that too many players, relative to the minimax predictions, exhibit statistically significant serial dependence in the runs tests (Table XVI); in fact there are about three times the number of rejections observed for professional players.

TABLE XII

RUNS TESTS IN O'NEILL'S EXPERIMENT WITH PROFESSIONAL PLAYERS[a]

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

[a]The ** and * denote rejections at the 5 and 10 percent levels, respectively.

TABLE XIII

RELATIVE FREQUENCIES OF CARD CHOICES AND WIN PERCENTAGES IN O'NEILL'S
EXPERIMENT WITH COLLEGE STUDENTS[a]

A. Frequencies

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

[a]In Panel A the numbers in parentheses represent minimax predicted relative frequencies, whereas those 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 standard deviation of the observed row player mean percentage win under the minimax hypothesis.

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 XIII in the next section for a comparison), the minimax model continues to be rejected. Given that the circumstances of the experiment are identical for college students and professional players, the results indicate that field skills and experience are indeed important.

TABLE XIV

RELATIVE FREQUENCIES OF CARD CHOICES IN O'NEILL'S EXPERIMENT WITH COLLEGE STUDENTS AND MINIMAX MULTINOMIAL TESTS[a]

| | Row Player (R) Choices | | | | Column Player (C) Choices | | | | p-Values for Tests of Minimax Multinomial Models | | |
| | | | | | | | | | Row | Column | Both |
Pair #	1	2	3	J	1	2	3	J	Player	Player	Players
1	0.225	0.270**	0.195	0.310**	0.140**	0.205	0.260**	0.395	0.022‡	0.065†	0.054†
2	0.205	0.180	0.160	0.455*	0.185	0.205	0.210	0.400	0.316	0.950	0.916
3	0.200	0.205	0.215	0.380	0.230	0.190	0.225	0.355	0.930	0.451	0.624
4	0.145**	0.215	0.155*	0.485**	0.175	0.180	0.225	0.420	0.031‡	0.604	0.011‡
5	0.135**	0.190	0.235	0.440	0.195	0.175	0.195	0.435	0.096†	0.732	0.028‡
6	0.185	0.215	0.235	0.365	0.215	0.230	0.230	0.325**	0.515	0.184	0.048‡
7	0.230	0.185	0.215	0.370	0.200	0.150*	0.165	0.485**	0.615	0.062†	0.078†
8	0.195	0.225	0.165	0.415	0.185	0.225	0.185	0.405	0.575	0.780	0.024‡
9	0.150*	0.215	0.200	0.435	0.200	0.215	0.195	0.390	0.342	0.960	0.085†
10	0.280**	0.260**	0.200	0.260**	0.250*	0.185	0.210	0.355	0.000‡	0.280	0.000‡
11	0.195	0.175	0.180	0.450	0.225	0.260**	0.205	0.310**	0.513	0.040‡	0.203
12	0.280**	0.210	0.180	0.330**	0.215	0.220	0.175	0.390	0.025‡	0.729	0.000‡
13	0.175	0.195	0.195	0.435	0.200	0.200	0.210	0.390	0.732	0.985	0.998
14	0.170	0.230	0.260**	0.340*	0.195	0.195	0.205	0.405	0.066†	0.993	0.160
15	0.140**	0.210	0.200	0.450	0.195	0.205	0.200	0.400	0.175	0.997	0.677
16	0.245*	0.195	0.190	0.370*	0.225	0.215	0.160	0.400	0.457	0.484	0.020‡
17	0.195	0.160	0.200	0.445	0.210	0.205	0.200	0.385	0.451	0.971	0.514
18	0.265**	0.210	0.185	0.340*	0.215	0.205	0.180	0.400	0.096†	0.885	0.036‡
19	0.300**	0.185	0.190	0.325**	0.255**	0.240	0.195	0.310**	0.004‡	0.034‡	0.000‡
20	0.195	0.205	0.165	0.435	0.215	0.195	0.205	0.385	0.596	0.943	0.744

[a]The ** and * denote rejection of the minimax binomial model for a given card and player at the 5 and 10 percent levels, respectively. Similarly, ‡ and † denote rejection at those levels of the minimax multinomial model based on Pearson statistic and a $\chi^2(3)$ for all cards chosen by a given row or column player, and based on Pearson statistic and a $\chi^2(15)$ for all cards chosen by both players.

TABLE XV

TESTING THAT COLLEGE STUDENTS EQUATE THEIR STRATEGY PAYOFFS TO THE EQUILIBRIUM
RATES IN O'NEILL'S EXPERIMENT[a]

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

[a]The ** and * denote rejections at the 5 and 10 percent levels, respectively; ‡ and † denote rejections at the same
levels when the four cards (1, 2, 3, J) are treated on an individual basis.

TABLE XVI

RUNS TESTS IN O'NEILL'S EXPERIMENT WITH COLLEGE STUDENTS[a]

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

[a]The ** and * denote rejections at the 5 and 10 percent levels, respectively.

4. DISCUSSION

4.1. *Evaluation of the Results*

Over the past decades, many experimental attempts have been made to test the theory of minimax, the results of which seem to go against the theory.[24] The first tests that yielded some support for equilibrium theory used empirical data obtained from competitive sports. Walker and Wooders (2001), for instance, brought initialevidence showing that top tennis players behave much more closely to equilibrium predictions than experimental subjects whose play had been observed thus far. Palacios-Huerta (2003) offered additional empirical evidence showing that professional soccer players' play is consistent with minimax theory. To properly evaluate our experimental results and compare them with the previous literature, let us describe some implications of minimax theory and the circumstances under which they are testable.

When players behave according to the unique mixed-strategy equilibrium of a given game, we have the following implications:

    1. The distribution of play should be close to the equilibrium one.

    2. Marginal frequencies should be close to the equilibrium mixtures.

    3. For each player, success rates should be (a) equal across chosen actions, and (b) equal to the equilibrium success rate.

    4. There should be no contemporaneous correlation.

    5. There should be no serial correlation.

These are direct implications of equilibrium play. There are other, second order, implications as well:

    6. When one tests for implication 2, the distribution of the resulting *p*-values across individuals and pairs should be close to uniform. Similarly, in the tests of implication 3, the distribution of *p*-values across individuals should be close to uniform.

Not all the above implications are equally strong. For example, implication 1 implies 2, but not the other way around. Indeed, marginal frequencies may well be the equilibrium ones, but due to contemporaneous correlation, the distribution of play may be far from the equilibrium distribution. Similarly, implication 1 implies 3, but not the other way around. Also, none of 1, 2, and 3, either jointly or separately, implies 6, and clearly 4 does not imply 5. Needless to say, the failure of even a single one of the above implications is enough to conclude that the observed behavior is not consistent with equilibrium. Nevertheless, if we are willing to go beyond the reject/no reject approach of classical hypothesis testing, we may say that the more implications are not rejected and the stronger these are, the more closely the observed behavior conforms to the theory.[25]

---

[24]See O'Neill (1987), Erev and Roth (1998), and the references therein.

[25]See O'Neill (1991) for an argument supporting a Bayesian approach to hypothesis testing combined with a measure of closeness of the results to the predictions that is much less rigid than

When the game is not known, as in the case of empirical studies such as Walker and Wooders (2001) and Palacios-Huerta (2003), implications 1, 2, and 3(b) cannot be tested. This is because since the game is not known, the equilibrium is not known either. On the other hand, implication 3(a) can be tested (using the maximum likelihood estimates of the success rates) even when the game is not known. Further, one can use the resulting p-values to check the corresponding implication 6. The favorable evidence reported in Walker and Wooders (2001) is based only on the nonrejection of implication 3(a) and implication 6 when applied to 3(a). Similarly, the favorable evidence reported in Palacios-Huerta (2003) is based on the nonrejection of implications 3(a), 6 applied to 3(a), and, in addition, implication 5.

When the game is known, as is typically the case in experimental studies like the present one, all the above implications are testable. In other words, the number and strength of the implications we can test is greater than in the field studies where the theory found support. Our qualitative results for the case of professional soccer players can be summarized as follows:

- Implications 1 and 2 are not rejected at the individual level, both in the penalty kick game and in O'Neill's game. At the aggregate level, these implications are not rejected for O'Neill's game.
- Although in the penalty kick game implication 1 is rejected at the aggregate level (and so is implication 2 for row players), an alternative hypothesis that the row players play the mixture 1/3-2/3 (only 3 percentage points apart from the theoretical 0.36-0.63 mixture), and the column players play the equilibrium mixture is far from being rejected. This alternative hypothesis is consistent with implication 4 of no contemporaneous correlation, and although 1/3 is statistically significantly different from 0.36, it seems it would be unfair to blame players for not hitting 0.36 precisely, especially when the opponents, even if they were sharp enough to notice it, could barely increase their success rate.
- Implications 3 and 4 are not rejected.
- There is little evidence of serial correlation.
- In O'Neill's game, implication 6 is rejected when applied to implication 2 at the player level but not at the pair level. It is also rejected when applied to implication 3(b) but not to implication 3(a).[26]

Based on the above, we can say that although strictly speaking minimax theory is rejected, it is still useful in explaining the experimental behavior of professional soccer players in the lab. More importantly perhaps, as the testable implications in the field (3(a), 5, and 6 applied to 3(a)) find support in the field and also find strong support in the lab both in the penalty kick and O'Neill's

---

the reject/no reject dichotomy. He argues that this approach may be particularly appropriate for situations which make precise point predictions like in the games we study.

[26]The p-values of the corresponding Kolmogorov–Smirnov tests are 0.267 for implication 3(a) and 0.002 for implication 3(b).

games, the evidence in the lab represents a good predictor for behavior in the field.

Last, with regard to the main evidence against the theory, it consists of the fact that in O'Neill's game the marginal frequencies of one-third of the players' action choices are too close to the equilibrium mixtures to have been generated by an i.i.d. process. Nonetheless, as long as players' behavior is largely unpredictable to other players, which seems to be the case given that the data pass our tests of serial independence, we may safely say that the minimax theory does well in explaining our soccer players' choices.

4.2. *Are All Professionals Perfect i.i.d. Randomizers?*

The time series evidence from the runs tests is not inconsistent with this hypothesis in any of the two experiments we have implemented. Yet, the cross-sectional evidence in O'Neill's game showed that about one-third of the players in this experiment chose actions with a relative frequency that is too close to the equilibrium distribution. While this is evidence that minimax theory is useful for predicting subjects' behavior, it is also evidence that these players are not choosing their actions according to an exact i.i.d. process. Next, we examine whether similar evidence of this lack of randomization may also be present in the 2×2 game.

Recall that in the 2×2, penalty kick game, the hypothesis that professionals play according to the equilibrium was rejected at the aggregate level. The reason is that the aggregate relative frequencies with which row players choose their actions are statistically different from the equilibrium mixtures. This suggests the hypothesis that subjects do randomize, but with nonequilibrium probabilities. Consider the following hypothesis: "At each stage, row players choose L and R with probabilities 1/3 and 2/3, respectively, and column players choose with probabilities 0.462 and 0.538."[27] It turns out that a chi-squared test based on Pearson goodness of fit indicates that this new hypothesis cannot be rejected at conventional significance levels since it yields a statistic of 0.296 whose p-value is 0.96. Further, we can perform for each player a chi-squared test of conformity with our new hypothesis and check whether the resulting distribution of p-values is significantly different from uniform $U[0, 1]$. Indeed, it turns out that a Kolmogorov–Smirnov test rejects the hypothesis that the p-values are draws from a uniform distribution; it yields a statistic of 1.99 whose p-value is 0.0007. The reason is that, as in O'Neill's experiment, the p-values are much higher than expected. In other words, our alternative theory predicts subjects' choices "too well."

The excessive closeness of the observed frequencies to the hypothesized ones suggests that subjects do not randomize, but rather *try* to "match" some probabilities they have, consciously or unconsciously, in mind. In the case of the

---

[27]These are the empirical frequencies. If we replace column players' frequencies by the equilibrium ones, the qualitative results do not change.

penalty kick game, it does not seem unreasonable that row players may try to keep their choices of *L* and *R* close to the 1/3-2/3 distribution. This distribution is very close to the equilibrium one; the difference is not likely to be detected by the opponent and it is arguably much simpler to play than the equilibrium distribution. This may explain why they seem to try to choose *R twice as often* as *L*. Similarly, in O'Neill's game about one-third of the subjects seem to try to keep their choices too close to the equilibrium distribution, which, interestingly enough, also prescribes choosing the *J* card *twice as often* as any other given card.

4.3. *Features that Catalyze Equilibrium Play*

There is a growing literature in experimental economics that shows how certain anomalies found in experiments with students as subjects are attenuated when the experiments are implemented, instead, with a nonstandard pool of subjects in their own natural environment (e.g., List (2003, 2004)). On the other hand, there is also evidence that experienced players still exhibit some nonequilibrium behavior when they play under laboratory conditions where they perform abstract tasks that lack a familiar context and that may not capture all the relevant aspects of the environment encountered in the field.[28] The nonequilibrium behavior observed in some professionals in O'Neill's game may, in part, be attributed to this factor. Recall that this game is unfamiliar to them and that the equilibrium requires treating a subset of available choices (the three non-*J* cards) identically to each other. Clearly, this component of the equilibrium represents an abstract task that professionals have never encountered in the field. Indeed, if we look only at the distribution of non-*J* card choices and compare it to the equilibrium distribution 1/3-1/3-1/3, we find that the cross-sectional distribution of the *p*-values of the tests is even more skewed than the distribution in the tests for all the cards. Thus, subjects exhibit a tendency to match particularly well the equilibrium frequencies of these three cards.[29] While this may not be particularly surprising since the written instructions of the experiment already treat the *NJ* cards identically, this evidence

---

[28]Kagel (1995), for instance, suggests that the reasons why professional bidders from the construction industry fall prey to the winner's curse in the experiments he cites are that (i) experiments strip away a number of contextual cues that they employ in field settings and (ii) the bidding environment created in the experiment is not representative of the environment encountered in the field, as the construction industry has private value and repeated play elements that are not present in the experiments.

[29]Matching frequencies is a behavior consistent with the evidence in the economics and psychology literature documenting that agents exhibit a law of small numbers bias (see, e.g., Tversky and Kahneman (1971), Rabin (2002), and other references therein). Although it is the time series properties (negative autocorrelation) of this law that have been the focus of the literature, the cross-sectional implication is also quite apparent: a distribution of *p*-values skewed toward 1.

indicates that the unfamiliar component of the equilibrium in this game drives the one departure from minimax that some subjects exhibit.[30]

Last, there is also a literature that addresses the effect of context as an inducement toward equilibrium behavior. Cooper, Kagel, Lo, and Gu (1999), for instance, showed that context speeds up the learning process of Chinese managers toward equilibrium play in a signaling game. In the psychology literature, Cosmides (1989) and Gigerenzer and Hug (1992) reported experimental evidence that the introduction of context dramatically reduces anomalous behavior in the unfamiliar but arguably simple Wason selection task.[31] In our experiments, professional soccer players were given no contextual cues. Although our 2×2 games were calibrated to be as close as possible to the real life penalty kick, players were not informed of this fact. Furthermore, O'Neill's 4×4 game is far from being representative of any interactive situation soccer players may encounter in their everyday life. Therefore, it seems somewhat striking that players exhibited very few departures from minimax play. The following observations may help elucidate our results:

SIMPLICITY OF THE GAMES: The zero-sum games studied in our experiments are very simple and do not require too much abstraction. They are complete information games with very few actions. In fact, our 2×2 game requires little more than performing the routine, day-to-day tasks of professional players. Note that besides playing one official game per week, these people practice 4–5 hours a day, 4–5 days a week, for 10 months a year, and soccer is a game that involves many zero-sum strategic situations (not only penalty kicks) that require randomization.[32] In this sense, our subjects might have spent a large part of their lives attempting to generate i.i.d. sequences while facing opponents who do the same. O'Neill's game on the other hand is simple in the sense that payoffs involve no probabilities and there are only two payoff levels; all that players need to know is that there are outcomes in which they win and outcomes in which they lose. Compare this simplicity with the informational requirements of common value auctions like the ones used in several experimental studies (e.g., Kagel (1995), Harrison and List (2008)). In these type

[30]The strategic symmetry of the non-J cards should be as obvious to students as it is to professionals. Consistent with this hypothesis, while the Kolmogorov–Smirnov test does not reject the hypothesis that the p-values are drawn from a uniform distribution, we do find 11 subjects whose p-values are above 0.85, when the expected number of such players is only 6. Thus, these players seem to be matching too well the equilibrium frequencies. We are indebted to a referee for raising this point.

[31]Ortmann and Gigerenzer (2000) summarized the results of the effect of context on experimental outcomes in this task.

[32]Data from the weekly official games indicate that the goalkeepers in our sample are involved, on average, in twelve penalty kicks per year, and our kickers are involved in roughly seven kicks per year. About one-third of the kickers can be considered "designated penalty kick takers" for their teams. For this subset, the average number of penalty kicks is eleven. No data are available from their day-to-day practices.

of auctions, there is an underlying (possibly infinite) set of states of the world over which each player holds a prior, and each player has a state dependent utility function. Unlike in O'Neill's game, both the priors and the utility functions affect the equilibria of these Bayesian games. Strategies in these games are very complicated objects, and the equilibrium ones are much more difficult to calculate, or even guess, than those in our zero-sum games.

UNIQUENESS OF THE EQUILIBRIUM: Both our zero-sum games have unique equilibria that are independent of the risk preferences of the players. Players only need to know that both players want to win. Even the simple signaling games of Cooper, Kagel, Lo, and Gu (1999) are an order of magnitude more complex than our zero-sum games.[33] They usually have several equilibria which even the most accepted refinements fail to disqualify. It is not completely surprising then that some contextual cues may help players coordinate on one of them. However, there is not much to coordinate on in our zero-sum games. Moreover, we know of no solution concept that does not select the minimax strategies in zero-sum games. Thus, it seems that context does not add much to the structure of the simplest of these games.

These considerations may help to explain why none of the potential drawbacks and limitations associated with the artificial environment that represents a laboratory seems to induce professional players to play very differently from the way they play in the field (in the penalty kick experiment) and from minimax equilibrium play (in both games). At the same time, they are not sufficient for inexperienced student subjects to play according to the predictions of equilibrium.

5. ADDITIONAL EVIDENCE AND EXTENSIONS

In this section we discuss some additional evidence we have collected and studied.

TESTS OF SERIAL INDEPENDENCE: We have considered a logit model for individual players to study whether own and opponent's past choices and outcomes, alone and interactive, play a role in determining current choices. For O'Neill's experiment, we have followed the formulation that Brown and Rosenthal (1990) study, whereas for the penalty kick experiment, we have followed the formulation suggested in Slonim, Roth, and Erev (2003). 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

---

[33]This may explain the observation in Cooper, Kagel, Lo, and Gu (1999) that in abstract tasks without contextual cues, students tend to perform better than old managers with inferior education.

is rejected for very few professional subjects, especially in the penalty kick experiment, whereas it is rejected for up to six times many more students.

We have also pooled all forty subjects for each experiment and class of players, and estimated the binary choice dynamic panel data model with predetermined endogenous variables and unobserved individual heterogeneity developed in Arellano and Carrasco (2003). The model controls for the state dependence that may be caused by past choices and past outcomes. The results confirm previous findings that no lagged endogenous variables (past own and opponent's choices and outcomes alone or interacted) are significant for professionals. Negative autocorrelation and positive reinforcement, however, significantly characterize the behavior of students.

ROBUSTNESS: We have also studied some variations in the experiments with smaller samples of subjects. For instance, in the penalty kick experiment we used payoffs that are entirely different from the scoring probabilities occurring in the field. We 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 much smaller, we find that none of these modifications of the experimental procedures seems to cause any significant changes in the basic results obtained earlier: professional soccer players continue to play with a high degree of accuracy with respect to the equilibrium predictions, while college students do not.

EXTENSIONS: 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 male 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. Playing in these leagues is still quite competitive. Amateur teams practice as often as professional ones and have exactly the same 10-month playing schedule. Also, 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 implemented both the penalty kick experiment and O'Neill's experiment for these subjects. Tables XVII and XVIII report the main results along with those obtained with professional players and college students with no soccer experience presented earlier. In Table XVIII, in addition, we include the results of Pearson's tests of equality of winning rates when, rather than using

TABLE XVII

SUMMARY STATISTICS IN PENALTY KICK EXPERIMENT

	Choice	Equilibrium	Professional Soccer Players	Students with Soccer Experience	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					
Row player (all cards)		1 (2)	0 (1)	1 (3)	2 (3)
Column player (all cards)		1 (2)	1 (2)	2 (2)	3 (10)
Both players (all cards)		1 (2)	1 (1)	1 (3)	3 (9)
All cards		4 (8)	4 (7)	9 (12)	12 (20)
III. Equality of Win Rates					
Rejections at 5 (10) percent		2 (4)	0 (2)	2 (3)	5 (12)
Aggregate Pearson statistic		—	31.60	42.45	65.69
p-value		—	0.826	0.366	0.006
IV. Runs Tests for 40 Players					
Rejections at 5 (10) percent		2 (4)	2 (4)	3 (5)	7 (13)

the equilibrium value, we use its maximum likelihood estimate,[34] as well as the original results of O'Neill's experiment reported in Brown and Rosenthal (1990) and Walker and Wooders (2001).

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 field skills and years of 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.

6. CONCLUDING REMARKS

This paper exploits three distinct features of soccer: (i) the existence of a precisely defined strategic situation played in the soccer field whose formal structure can be reproduced in the laboratory, (ii) the fact that this situation has

---

[34]This is the procedure that Walker and Wooders (2001) followed 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.

TABLE XVIII

SUMMARY STATISTICS IN O'NEILL'S EXPERIMENT

	Card	Equilib-rium	Professional Soccer Players	Students with Soccer Experience	Students with No Soccer Experience	Students in O'Neill[a]
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 (std. deviation)		0.400 (0.007)	0.394	0.403	0.391	0.410
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. Equality of Success Rates Across Strategies and to the Equilibrium Rate Using NJ and J[b]						
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	6.78×10^{-17}
All column players p-value		—	0.654	0.959	0.0042	1.90×10^{-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	4.93×10^{-8}
All column players p-value		—	0.298	0.387	0.002	1.45×10^{-8}
IV. Runs Tests						
Rejections at 5 (10) percent		2 (4)	2 (4)	3 (5)	7 (12)	15 (19)

[a]The results in this column 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 experiment is 1.25 greater than those reported in the first column, and the standard deviation for observed relative frequencies under minimax play in panel I is 0.009, rather than 0.007.

[b]In O'Neill's original experiment there are two pairs that represent extreme outliers. When these are ignored, the p-values remain very low (1.2×10^{-9} and 1.7×10^{-12}, respectively).

a unique mixed-strategy equilibrium, and (iii) the evidence that professional subjects play in a real life setting consistent with the equilibrium of this game. These features are helpful for designing a first artefactual field experiment about mixed-strategy interactions that isolates the role of laboratory context and that allows the comparison of field and laboratory behavior. We find that experience and skills may be transferred from the familiar soccer field to the highly unfamiliar laboratory when professional subjects play this game. Their skills are also valuable for playing very close to the equilibrium in a zero-sum game that subjects have never faced before. Some deviations from equilibrium in this game partially arise from a component that represents an abstract task that professionals have never encountered in the field.

These findings provide insights into the transfer of knowledge across strategic settings and also indicate that is important how experience is defined in experimental settings. Further, our results may have theoretical, methodological, and cognitive implications:

In terms of the theory, since professionals' behavior is very close to the equilibrium one, the theoretical concept of equilibrium 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 to which data that are typically used to inform game theory and other theoretical areas in various social sciences, as well, often come 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 seem largely applicable for predicting field behavior.

Last, from a cognitive perspective our findings are consistent with the idea that skills are learned unconsciously and are active in the solution of the games we have studied. In this sense, the "capacity of motivated subjects to find equilibrium outcomes … without cognitive awareness of this capacity" emphasized in Smith (2005) is supported, for the first time to our knowledge, in situations requiring use of mixed strategies. From this perspective, Camerer, Loewenstein, and Prelec (2005) discussed neurological evidence showing 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." 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. Similarly, we cannot disregard the idea that players who became professionals were born with greater aptitude for playing strategic zero-sum games. To delineate between treatment and selection effects, further treatments in the spirit of List (2003, 2006) would be needed.

Dept. of Economics, Brown University, Box B, 64 Waterman Street, Providence, RI 02912, U.S.A., and Dept. of Management, London School of Economics, London WC2A 2AE, U.K.; ipalacios@brown.edu

and

Dept. of Economics, Iowa State University, 260 Heady Hall, Ames, IA 50011, U.S.A., and Dept. of Economics, Ben-Gurion University, Beer-Sheva, 84105 Israel; ovolij@bgu.ac.il.

REFERENCES

ARELLANO, M., AND R. CARRASCO (2003): "Binary Choice Panel Data Models with Predetermined Variables," *Journal of Econometrics*, 115, 125–157. [109]

AUMANN, R. J. (1987): "Game Theory," in *The New Palgrave: A Dictionary of Economics*, Vol. 2, ed. by J. Eatwell, M. Milgate, and P. Newman. London: MacMillan Press, 460–482. [74]

BAR-HILLEL, M., AND W. A. WAGENAAR (1991): "The Perception of Randomness," *Advances in Applied Mathematics*, 12, 428–454. [84]

BINMORE, K., J. SWIERZBINSKI, AND C. PROULX (2001): "Does Minimax Work? An Experimental Study," *Economic Journal*, 111, 445–464. [74]

BROWN, J., AND R. ROSENTHAL (1990): "Testing the Minimax Hypothesis: A Re-Examination of O'Neill's Experiment," *Econometrica*, 58, 1065–1081. [73,79,97,108,110,111]

CAMERER, C. (1995): "Individual Decision Making," in *Handbook of Experimental Economics*, ed. by J. H. Hagel and A. E. Roth. Princeton, NJ: Princeton University Press, 587–703. [73,84]

——— (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton, NJ: Princeton University Press. [71,72]

CAMERER, C., G. LOEWENSTEIN, AND D. PRELEC (2005): "Neuroeconomics: How Neuroscience Can Inform Economics," *Journal of Economic Literature*, 43, 9–64. [112]

CHIAPPORI, P. A., S. D. LEVITT, AND T. GROSECLOSE (2002): "Testing Mixed Strategy Equilibrium when Players Are Heterogeneous: The Case of Penalty Kicks," *American Economic Review*, 92, 1138–1151. [73]

COOPER, D. J., J. H. KAGEL, W. LO, AND Q. L. GU (1999): "Gaming Against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers," *American Economic Review*, 89, 781–804. [107,108]

COSMIDES, L. (1989): "The Logic of Social Exchange: Has Natural Selection Shaped how Humans Reason? Studies with the Wason Selection Task," *Cognition*, 31, 198–276. [107]

DELOITTE & TOUCHE (2005): *Annual Review of Football Finance 2005*. London, Deloitte. [75]

EREV, I., AND A. E. ROTH (1998): "Predicting how People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, 88, 848–881. [103]

FIFA (2005): *Fédération Internationale de Football Association—Official Laws of the Game*. Chicago, IL: Triumph Books. [75,76]

GIBBONS, J., AND S. CHAKRABORTI (1992): *Nonparametric Statistical Inference*. New York, Marcel Dekker. [85]

GIGERENZER, G., AND K. HUG (1992): "Reasoning about Social Contracts: Cheating and Perspective Change," *Cognition*, 43, 127–171. [107]

GIGERENZER, G., AND P. M. TODD (1999): *Simple Heuristics that Make Us Smart*. New York: Oxford University Press. [74]

GILBOA, I., AND D. SCHMEIDLER (2001): *A Theory of Case-Based Decision*. Cambridge, U.K.: Cambridge University Press. [74]

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. [77]

HARRISON, G. W. (2005): "Field Experiments and Control," in *Research in Experimental Economics*, Vol. 10, ed. by J. Carpenter, G. W. Harrison, and J. A. List. Greenwich, CT: JAI Press, 17–50. [72]

HARRISON, G. W., AND J. A. LIST (2004): "Field Experiments," *Journal of Economic Literature*, 42, 1009–1055. [72]

———— (2008): "Naturally Occurring Markets and Exogenous Laboratory Experiments: A Case Study of the Winner's Curse," *Economic Jornal*, 118, forthcoming. [107]

HONG, J. T., AND C. R. PLOTT (1982): "Rate Filing Policies for Inland Water Transportation: An Experimental Approach," *Bell Journal of Economics*, 13, 1–19. [77]

HUME, D. (1739): *A Treatise of Human Nature* (Second Ed.). Oxford U.K.: Clarendon Press. [71,74]

KAGEL, J. H. (1995): "Auctions: A Survey of Experimental Research," in *Handbook of Experimental Economics*, ed. by J. H. Kagel and A. E. Roth. Princeton, NJ: Princeton University Press, 501–585. [106,107]

KAGEL, J. H., AND A. E. ROTH (1995): *Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press. [72]

LAZEAR, E., U. MALMENDIER, AND R. A. WEBER (2005): "Sorting in Experiments," Mimeo, Stanford University. [72]

LIST, J. A. (2003): "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, 118, 41–71. [106,112]

———— (2004): "Neoclassical Theory versus Prospect Theory: Evidence from the Marketplace," *Econometrica*, 72, 615–625. [106]

———— (2006): "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions," *Journal of Political Economy*, 114, 1–37. [112]

MARCA (2005): *Guia de la Liga 2006*. Madrid: Recoletos Grupo de Comunicación. [75]

MOOKHERJEE, D., AND B. SOPHER (1994): "Learning Behavior in an Experimental Matching Pennies Game," *Games and Economic Behavior*, 7, 62–91. [84]

NEURINGER, A. J. (1986): "Can People Behave "Randomly"?: The Role of Feedback," *Journal of Experimental Psychology: General*, 115, 62–75. [84]

———— (2002): "Operant Variability: Evidence, Function, and Theory," *Psychonomic Bulletin and Review*, 9, 672–705. [73,84]

O'NEILL, B. (1987): "Nonmetric Test of the Minimax Theory of Two-Person Zerosum Games," *Proceedings of the National Academy of Sciences*, 84, 2106–2109. [73,77-79,92,103,111]

———— (1991): "Comments on Brown and Rosenthal's Reexamination," *Econometrica*, 59, 503–507. [103]

ORTMANN, A., AND G. GIGERENZER (2000): "Reasoning in Economics and Psychology: Why Social Context Matters," in *Cognition, Rationality and Institutions*, ed. by M. E. Streit, U. Mummert, and D. Kiwit. Berlin: Springer, 131–151. [107]

PALACIOS-HUERTA, I. (2003): "Professionals Play Minimax," *Review of Economic Studies*, 70, 395–415. [73,77,103,104]

PALACIOS-HUERTA, I., AND O. VOLIJ (2008): "Supplement to 'Experientia Docet: Professionals Play Minimax in Laboratory Experiments'," *Econometrica Supplementary Material*, 76, http://www.econometricsociety.org/ecta/Supmat/6315_instructionstoexperimentalsubjects.pdf. [79]

RABIN, M. (2002): "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics*, 117, 775–816. [106]

RAPOPORT, A., AND R. B. BOEBEL (1992): "Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis," *Games and Economic Behavior*, 4, 261–283. [84]

RAPOPORT, A., AND D. V. BUDESCU (1992): "Generation of Random Sequences in Two-Person Strictly Competitive Games," *Journal of Experimental Psychology: General*, 121, 352–363. [84]

RUBINSTEIN, A. (2007): "Instinctive and Cognitive Reasoning: A Study of Response Times," *Economic Journal*, 117, 1243–1259. [78]

SELTEN, R. (1998): "Features of Experimentally Observed Bounded Rationality," *European Economic Review*, 42, 413–436. [74]

SHACHAT, J. M. (2002): "Mixed Strategy Play and the Minimax Hypothesis," *Journal of Economic Theory*, 104, 189–226. [73,79,97]

SIMON, H. A. (1983): *Reason in Human Affairs*. Stanford, CA: Stanford University Press. [74]

SLONIM, R. L., A. E. ROTH, AND I. EREV (2003): "Regularities in Play of Zero-Sum Games with Unique Mixed-Strategy Equilibria," Mimeo, Case Western Reserve University. [84,108]

SMITH, V. L. (2005): "Behavioral Economics Research and the Foundations of Economics," *Journal of Socio-Economics*, 34, 135–150. [112]

TVERSKY, A., AND D. KAHNEMAN (1971): "The Belief in the Law of Small Numbers," *Psychological Bulletin*, 76, 105–110. [84,106]

WALKER, M., AND J. WOODERS (2001): "Minimax Play at Wimbledon," *American Economic Review*, 91, 1521–1538. [73,85,95,97,103,104,110,111]

WEIBULL, J. (2004): "Testing Game Theory," in *Advances in Understanding Strategic Behaviour: Game Theory, Experiments and Bounded Rationality. Essay in Honour of Werner Güth*, ed. by S. Huck. New York: Palgrave Macmillan, 85–104. [72]

WOODERS, J., AND J. M. SHACHAT (2001): "On the Irrelevance of Risk Attitudes in Repeated Two-Outcome Games," *Games and Economic Behavior*, 34, 342–363. [76]