Review of Economic Studies (2003) **70**, 395–415 0034-6527/03/00150395$02.00

Professionals Play Minimax

IGNACIO PALACIOS-HUERTA
Brown University

First version received September 2001; *final version accepted October* 2002 (*Eds.*)

The implications of the Minimax theorem are tested using natural data. The tests use a unique data set from penalty kicks in professional soccer games. In this natural setting experts play a one-shot two-person zero-sum game. The results of the tests are remarkably consistent with equilibrium play in every respect: (i) winning probabilities are statistically identical across strategies for players; (ii) players' choices are serially independent. The tests have substantial power to distinguish equilibrium play from disequilibrium alternatives. These results represent the first time that both implications of von Neumann's Minimax theorem are supported under natural conditions.

1. INTRODUCTION

During the last few decades game theory has contributed to a reshaping of important aspects of the methodology of Economics and other social sciences. In a large part this is because the language, concepts and techniques of non-cooperative game theory have become central to many areas of the discipline. Undoubtedly, studying the interaction of ideally rational players greatly aids our understanding of the behaviour of real individuals, firms and other agents. Moreover, as Kreps (1991) remarks, "studying the former should aim at understanding the latter. The point of game theory is to help economists understand and predict what will happen in economic, social and political contexts".

Theoretical contributions should thus feed back to empirical analysis. However, testing the implications of the theory has proven extremely difficult in the literature. The primary reason is that many predictions often hinge on properties of the utility functions and the values of the rewards used. Even when predictions are invariant over classes of preferences, data on rewards are seldom available in natural settings. Moreover, there is often great difficulty in determining strategy sets and in measuring individuals' choices, effort levels, and the incentive structures they face. As a result, even the most fundamental predictions of game-theoretical models have not yet been supported empirically in real situations.

In view of the substantial problems associated with testing theoretical predictions using natural data, many authors have been compelled to test them in experimental settings. Interestingly enough, despite the controlled structure of experiments, the results of many experiments during the last few decades typically reject the assumption that subjects are playing according to the theoretical implications of *equilibrium* play. For instance, a number of experiments have evaluated the empirical validity of von Neumann's Minimax Theory for two-person zero-sum games, a theory that occupies a central position in our understanding of strategic situations. The results have been mixed and are often rather negative (see, for instance, Brown and Rosenthal (1990), Rapoport and Boebel (1992), Camerer (2003), and the references therein).

It is rightly argued, however, that in experimental settings individuals are often exposed to games and situations that they have not faced previously and that, despite their usual simplicity, it may not be possible for them to become very proficient in the limited timeframe of an experiment. This in turn has contributed to generating an important literature dealing with out-of-equilibrium play, or "learning", in experimental games.[1] Consequently, the state of affairs is such that even though the ability of game theoretical models to explain the diversity of individual behaviour in socioeconomic environments has significant theoretical and practical implications, the major equilibrium predictions of the theory have received little or no empirical support.

This paper offers an empirical examination of an aspect of the theory of strategic interactions. A fundamental concept in many strategic situations is that one's actions must not be predictable by one's opponent. The theory of mixed strategy play, including von Neumann's Minimax theorem and the general notion of a Nash equilibrium in mixed strategies, remains the essential basis of our theoretical understanding of strategic situations that require unpredictability.[2] In this paper we use data from a natural strategic play in professional sports to provide an empirical test of the Minimax theorem, an approach that has been recently followed by Walker and Wooders (2001). Moreover, the specific features of the play and its environment allow the analysis to overcome the usual difficulties that have plagued previous empirical attempts in the literature, both in natural and experimental settings.

In professional sports participants are experts at their games. This clearly presents a notable advantage over many experiments, as the individuals are already the most proficient agents of the situations of the game. However, despite the fact that situations that require unpredictable play and mutual outguessing are pervasive in sports, it is normally not feasible to conduct accurate empirical tests of the Minimax hypothesis in these scenarios. The reason is that players generally have many strategies available and outcomes are numerous. In addition, outcomes are generally not decided immediately after players choose their initial actions. There are subsequent strategic choices that usually play a crucial role in determining final outcomes. As a result, it is essentially impossible to obtain detailed data on all relevant variables to conduct informative statistical tests.

All of these usual difficulties and drawbacks are overcome in the natural setting we examine in this paper. We focus on a specific one-shot two-person zero-sum game in professional sport that requires unpredictability and mutual outguessing. As in the typical experimental setting, the game has a precisely defined set of rules, few strategies are available, outcomes are decided immediately after strategies are chosen, and all relevant information is observable. In contrast with most experiments, however, professional players are highly motivated and experts at the game. The analysis is concerned with a play of the world's most popular game: penalty kicks in soccer. To test the implications of the Minimax theorem, we exploit a unique data set of more than 1400 penalty kicks in professional soccer games that includes very detailed information on all relevant aspects of the play, especially actions and outcomes.

As a brief summary of the main results, we find that, as predicted by the theory of mixed strategy play: (i) winning probabilities are statistically identical across strategies, and (ii) players' choices are independent draws from a random process. The first result has been notably difficult to obtain in the literature. As to the second, essentially all previous tests of randomness in experimental research in psychology and economics have found that individuals "switch strategies" too often to be consistent with random play. When individuals are asked to generate or identify random sequences these sequences often have negative autocorrelations (see Bar-Hillel and Wagenaar (1991), Camerer (1995) for reviews of the literature). In our tests of the Minimax hypothesis, professional players are found to be capable of behaving perfectly randomly. Their sequences neither exhibit negative or positive autocorrelation, and choices do not depend on one's own previous play, on the opponent's previous plays or on past outcomes.

1. See Erev and Roth (1998), Camerer and Ho (1999), Stahl (2000), and many references therein.
2. Osborne and Rubinstein (1994) discuss a number of interpretations of mixed strategy equilibrium.

We take these two results as consistent with the implications of the Minimax theorem. In this sense, and to the best of our knowledge, they represent the first time that the fundamental notion of Nash equilibrium in mixed strategies is supported with real data.

The remainder of the paper is organized as follows. The following section describes the structure and setting of the play, and the empirical hypothesis that will be evaluated. Section 3 describes the data. Section 4 is devoted to the empirical analysis. Section 5 contains an evaluation of the power of the tests against non-Minimax behaviour, and a discussion of some extensions of the empirical analysis. Finally, Section 6 concludes.

2. MIXED STRATEGIES IN PENALTY KICKS

In soccer, a penalty kick is awarded against a team which commits one of the ten punishable offences inside its own penalty area while the ball is in play. The world governing body of soccer, the *Fédération Internationale de Football Association* (FIFA), describes in detail the simple rules that govern this play in the *Official Laws of the Game* (FIFA, 2000).[3] First, the position of the ball and the players are determined as follows:

- "The ball is placed on the penalty mark in the penalty area.[4]
- The player taking the penalty kick is properly identified.
- The defending goalkeeper remains on the goal line, facing the kicker, between the goalposts, until the ball has been kicked.
- The players other than the kicker are located inside the field of play, outside the penalty area, behind the penalty mark, and at least 10 yards (9·15 m) from the penalty mark."

  The procedure to be followed is described as follows:
- "The player taking the penalty kicks the ball forward.
- He does not play the ball a second time until it has touched another player.
- A goal may be scored directly from a penalty kick."

Each penalty kick involves two players: a kicker and a goalkeeper. In the typical kick the ball takes about 0·3 s to travel the distance between the penalty mark and the goal line; that is, it takes less than the reaction time plus goalkeeper's movement time to possible paths of the ball. Hence, both kicker and goalkeeper must move simultaneously.[5] The penalty kick has only two possible outcomes: score or no score. Players have few strategies available and their actions are observable. The spin of the kick plays no role. There are no second penalties in case a goal is not scored. The initial location of both the ball and the goalkeeper is always the same: the ball is

3. To William McCrum belongs the credit of inventing the original penalty kick rule. As a member of the Irish Football Association he submitted the penalty kick rule to the International Football Board in 1891. The rule was passed unanimously.

4. Law 1 is concerned with the dimensions of the field of play: "the penalty area is defined at each end of the field as follows: two lines are drawn at right angles to the goal line, 18 yards (16·5 m) from the inside of each goalpost. These lines extend into the field of play for a distance of 18 yards (16·5 m) and are joined by a line drawn parallel with the goal line. The area bounded by these lines and the goal line is the penalty area. Within each penalty area a penalty mark is made 12 yards (11 m) from the midpoint between the goalposts and equidistant to them. . . . Goals must be placed on the centre of each goal line. They consist of two upright posts equidistant from the corner flagposts and joined at the top by a horizontal crossbar. The distance between the posts is 8 yards (7·32 m) and the distance from the lower edge of the crossbar to the ground is 8 feet (2·44 m)" (FIFA, 2000).

5. Miller (1998) reports evidence on ball speed, reaction times, and movement times from all the penalty kicks in four World Cups which confirms that, as intended by the rule, goalkeepers must optimally begin their movement at the point of foot–ball contact and that kickers must choose their kicking side before goalkeepers move.

placed on the penalty mark and the goalkeeper positions himself on the goal line, equidistant to the goalposts.[6] Finally, the outcome is decided immediately after players choose their strategies.

The clarity of the rules and the detailed structure of this simultaneous one-shot play resembles those in the theoretical setting and in many experiments. In this sense, it presents notable advantages over other plays in professional sports and other natural settings. In baseball, pitchers and batters have many actions available and possible outcomes are numerous. In cricket and tennis, possible outcomes are limited but players have many strategic choices available. Even in serves, the direction of the serve, its spin, and the initial location of the opponent are all important strategic choices. Another difference is that outcomes are typically not decided immediately. After a player serves, there is subsequent strategic play that often plays a crucial role in determining the final outcome. Notable difficulties clearly arise both in modelling such situations theoretically and, especially, in observing all strategic choices in a given play. In an original paper, however, Walker and Wooders (2001) examine whether the choice of first serves alone by professional tennis players is consistent with equilibrium play. Although no other strategic variables are considered, second serves are ignored, and outcomes are not decided immediately, their findings support the implication that winning probabilities are statistically identical across strategies. Players, however, switch serving strategies too often to be consistent with random play, and hence with one implication of the Minimax theorem.

The characteristics of the play and the fact that it involves professional players—for whom arguably no learning is involved—allow the example in our case to overcome previous difficulties both in experimental and natural settings. They present a suitable opportunity for testing the Minimax hypothesis with natural data.

In what follows, we let the player's payoffs be the probabilities of success ("score" for the kicker and "no score" for the goalkeeper) in the penalty kick. The kicker wishes to maximize the expected probability of scoring, while the goalkeeper wishes to minimize it. Consider, for example, a simple 2×2 game-theoretical model of player's actions for the penalty kick and let π_{ij} denote the kicker's probabilities of scoring, where $i = \{L, R\}$ denotes the kicker's choice and $j = \{L, R\}$ the goalkeeper's choice, with $L =$ left, $R =$ right:

$i \backslash j$	L	R
L	π_{LL}	π_{LR}
R	π_{RL}	π_{RR}

This game has a unique Nash equilibrium when

$$\pi_{LR} > \pi_{LL} < \pi_{RL},$$
$$\pi_{RL} > \pi_{RR} < \pi_{LR}.$$

If the play in a penalty kick can be represented by this model, then equilibrium play requires each player to use a mixed strategy. Equilibrium theory then yields two sharp testable predictions about the behaviour of kickers and goalkeepers:

1. Success probabilities—the probability that a goal will be scored (not scored) for the kicker (goalkeeper)—should be the same across strategies for each player. Formally, let g_L denote the goalkeeper's probability of choosing left. This probability should be chosen so as to make the kicker's success probabilities identical across strategies. That is, g_L should satisfy $p_L^k = p_R^k$

6. The goalkeeper always chooses to be equidistant to the goalposts. The reason is that the distance between the goalposts is long enough (8 yards) so that he cannot afford to "invite" a professional kicker to shoot in a given direction. A professional kicker would score with an extremely high probability if the goalkeeper were not to locate himself 4 yards from each goalpost.

where

$$p_L^k = g_L \pi_{LL} + (1 - g_L)\pi_{LR},$$
$$p_R^k = g_L \pi_{RL} + (1 - g_L)\pi_{RR}.$$

Similarly, the kicker's probability of choosing left k_L should be chosen so as to make the goalkeeper's success probabilities identical across strategies, $p_L^g = p_R^g$, where

$$p_L^g = k_L(1 - \pi_{LL}) + (1 - k_L)(1 - \pi_{RL}),$$
$$p_R^g = k_L(1 - \pi_{LR}) + (1 - k_L)(1 - \pi_{RR}).$$

2. Each player's choices must be serially independent given constant payoffs across games. That is, individuals must be concerned only with instantaneous payoffs and intertemporal links between occurrences must be absent. Hence, players' choices must be independent draws from a random process; therefore, they should not depend on one's own previous play, on the opponent's previous play, on their interaction, or on any other previous actions.

We will test these fundamental predictions of von Neumann's Minimax theorem next. Before describing the data and implementing the empirical tests, two final observations are in order. First, it will be shown that there is a perfect symmetry between left-footed and right-footed kickers. In fact, the hypothesis that the game is identical for left-footed and right-footed kickers, up to the appropriate renaming of the actions, will not be rejected. Second, the distinct advantages and characteristics of this play make it suitable for empirical analyses other than the equilibrium implications for individual player behaviour implied by the Minimax theorem. Chiappori, Levitt and Groseclose (2002), for instance, are concerned with problems of aggregation across different players when few observations are available per player and individual level tests cannot be conducted. In particular, they examine a small data set of observations where individual players are involved in a limited number of penalty kicks (most are involved in five or fewer penalties). In situations where few observations are available per player it is not possible to test the Minimax implications for *individual* players. However, these authors show that it is still possible to test for various relevant cross-sectional implications that arise when aggregating observations across players using the available data.[7]

3. DATA

Data on 1417 penalty kicks have been collected during the period September 1995–June 2000 from professional games in Spain, Italy, England and other countries. The data come from the weekly TV programmes *English Soccer League* in the United States (Fox Sports World), *Estudio Estadio* in Spain (TVE), *Noventessimo Minuto* in Italy (RAI) and from various weekly programmes on the European station *Eurosport*. These programmes review all of the best plays in the professional games played during the week, including *all* penalty kicks that take place in the games.

The data include the names of the teams involved in the match, the date of the match, the names of the kicker and the goalkeeper for each penalty kick, the choices they take (left, centre, or right), the time at which the penalty kick is shot, the score at that time, and the final score in the match. They also include the kicker's kicking leg (left or right) and the outcome of the shot (goal or no goal).[8] More than 90% of all observations come from league matches in Italy,

7. The basic aggregate predictions of their analysis find support in their data set and are also substantiated in our data set.

8. The outcome "no goal" in the data includes in separate categories saves made by the goalkeeper and penalties shot wide, to the goalpost or to the crossbar by the kicker.

TABLE 1

Distribution of strategies and scoring rates

Score difference	#Obs.	LL	LC	LR	CL	CC	CR	RL	RC	RR	Scoring rate
0	580	16·9	1·3	21·0	4·3	0·8	5·6	19·4	0·6	27·9	81·9
1	235	19·1	0	19·1	4·2	0	2·5	28·0	0	26·8	77·8
−1	314	19·7	0·9	25·8	1·9	0	6·4	20·0	0·6	30·2	80·2
2	97	23·7	2·0	17·5	5·2	0	0	20·6	1·0	29·9	75·2
−2	114	26·3	0	25·4	3·5	0	3·5	16·6	0	24·5	78·0
3	27	14·8	0	18·5	3·7	0	11·1	22·2	0	29·6	77·7
−3	23	30·4	0	30·4	0	0	0	21·7	0	17·4	82·6
4	7	42·8	0	28·5	0	0	0	14·2	0	14·2	100
−4	12	25·0	0	25·0	0	0	16·6	16·6	0	16·6	83·3
Others	8	50·0	0	0	0	0	12·5	37·5	0	0	87·5
Penalties shot in:											
First half	558	21·1	0·8	19·8	3·9	0·3	3·5	20·0	0·3	29·7	82·9
Second half	859	18·7	0·9	23·2	3·3	0·3	3·6	22·8	0·5	26·3	78·3
Last 10 min	266	21·8	0	21·0	0·3	0	0·7	25·1	0	30·8	73·3
All penalties	1417	19·6	0·9	21·9	3·6	0·3	3·6	21·7	0·5	27·6	80·1
Scoring rate	80·1	55·2	100·0	94·2	94·1	50·0	82·3	96·4	100·0	71·1	

Note: The first letter of the strategy denotes the kicker's choice and the second the goalkeeper's choice. "R" denotes the R.H.S. of the goalkeeper, "L" denotes the L.H.S. of the goalkeeper, and "C" denotes centre.

Spain and England.[9] The leagues in these countries are considered to be the premier leagues in the world. The first two tables offer a basic description of the data.

Table 1 shows the relative proportions of the different choices made by the kicker and the goalkeeper (left (L), centre (C), or right (R)), with the total number of observations in the second left-most column. The first letter refers to the choice made by the kicker and the second to the choice made by the goalkeeper, both from the viewpoint of the goalkeeper. For instance, "RL" means that the kicker chooses to kick to the R.H.S. of the goalkeeper and the goalkeeper chooses to jump to his left. The right-most column shows the scoring rate for a given score difference. The term "score difference" is defined as the number of goals scored by the kicker's team minus the number of goals scored by the goalkeeper's team at the time the penalty is shot. For instance, a "−1" means that the kicker's team was behind by one goal at the time of the penalty kick.

The strategy followed by goalkeepers coincides with that followed by kickers in about half of all penalty kicks in the data set. Most are RR (27·6%), with 19·6% being LL and 0·3% being CC. Kickers kick to the centre relatively rarely (7·5% of all kicks), whereas goalkeepers choose C even less often (1·7%). The percentage of kicks where players' strategies do not coincide with each other are almost equally divided between LR (21·9%) and RL (21·7%). A goal is scored in 80·1% of all penalty kicks. The scoring rate is essentially 100% when the goalkeeper's choice does not coincide with the kicker's, and it is over 60% when it coincides. It is well known that soccer matches last two equal halves of 45 min, with a 15 min half-time interval. The scoring rate in the sample is slightly lower in the second half (78·3%) than in the first half (82·9%), and substantially lower in the last 10 min of a game (73·3%) than the overall average (80·1%). The average number of goals per match in the sample is 2·57. It is thus no surprise to observe that in most penalty kicks the score difference is 0, 1, or −1 at the time of the shot. For these score differences, the scoring rate is slightly greater in tied matches (81·9%), followed by the rate in

9. Most professional soccer matches take place in league tournaments, which typically last for a season of 9 months a year. All other observations come from cup competitions (elimination tournaments that are simultaneously played) and international games.

TABLE 2

Distribution of strategies and scoring rates by kicker type

Score difference	#Obs.	LL	LC	LR	CL	CC	CR	RL	RC	RR	Scoring rate
		Left-footed kickers									
0	174	17·8	1·7	20·1	6·3	0	8·6	22·9	0·5	21·8	82·7
1	73	28·7	0	30·1	4·1	0	2·7	19·1	0	15·0	78·0
−1	92	29·3	1·0	26·0	1·0	0	2·0	21·7	1·0	18·4	82·6
2	29	51·7	0	13·7	3·0	0	0	10·3	0	20·6	72·4
−2	30	40·0	0	13·3	3·0	0	3·0	20·0	0	20·0	76·6
All penalties	406	29·3	1·4	20·4	4·4	0	3·9	23·8	0	16·5	
Scoring rate	81·0	62·1	100	95·1	94·4	0	81·2	93·8	0	61·2	
		Right-footed kickers									
0	406	16·4	1·2	21·4	3·4	1·2	4·4	20·4	0·7	30·5	83·2
1	162	14·8	0	14·2	4·3	0	2·4	32·1	0	32·1	77·7
−1	222	15·7	1·0	25·6	2·2	0	0	19·3	1·0	35·1	80·6
2	68	11·7	2·9	19·1	5·8	0	0	25·0	1·4	33·8	76·4
−2	84	21·4	0	29·7	3·5	0	3·5	15·4	0	26·2	78·5
All penalties	1011	15·8	0·6	22·5	3·2	0·5	3·4	20·8	0·6	32·1	
Scoring rate	79·8	50·0	100	93·8	93·9	60·0	82·8	97·6	100	73·2	

Note: The first letter of the strategy denotes the kicker's choice and the second the goalkeeper's choice. "R" denotes the R.H.S. of the goalkeeper, "L" denotes the L.H.S. of the goalkeeper, and "C" denotes centre.

matches where the kicker's team is behind by one goal (80·2%), and then by the rate in matches where his team is ahead by one goal (77·8%).

Kickers may be classified into two types according to their kicking leg: left-footed and right-footed. Most kickers in the sample are right-footed, as is the case in the population of soccer players. Table 2 shows the distribution of strategies and scoring rates by kicker type and score difference.

As is clear, these two groups of kickers have different strong sides. Left-footed kickers shoot more often to the L.H.S. of the goalkeeper than to the R.H.S., whereas right-footed kickers shoot more often to the R.H.S. In the next section we will consider these to be their "natural sides" respectively, and choices will be renamed accordingly. Goalkeepers, in turn, tend to choose right more often than left when facing a right-footed kicker, and left more often than right when facing a left-footed kicker. Scoring rates are similar for the two player types for all penalties and for given score difference. Note also that kickers tend to be more successful when shooting to their natural sides. Other summary statistics, including evidence on Nash predicted frequencies and actual choices by kicker type, are reported in Palacios-Huerta (2002). We turn next to testing the empirical implications of mixed strategy equilibrium in this setting.

4. EMPIRICAL ANALYSIS

In the empirical tests of the implications of the Minimax theorem we start by considering the subjects that were involved in a relatively large amount of penalties. There are 22 kickers and 20 goalkeepers in the sample that were involved in at least 30 penalties each. All these players play in the Italian, Spanish and English professional leagues. Their identities are shown in the Appendix. For each of these players the observations in the data set include *all* the penalties they participated in during the period September 1995–June 2000, in the order that they took place.

Given that the roles are reversed for right-footed kickers and left-footed kickers, it would be erroneous to treat the games associated with these different types of kickers as equal. For this reason, in the remainder of this paper we will consider players' choices according to the kickers' natural sides. Whatever the kicker's strong foot, R denotes the "kicker's natural side" and L denotes the "kicker's non-natural side". When the kicker is right-footed the natural side R is the R.H.S. of the goalkeeper, and when the kicker is left-footed it is the L.H.S. of the goalkeeper. This means, for instance, that a left-footed kicker kicking to the goalkeeper's right is the same as a right-footed kicker kicking to the goalkeeper's left. Thus, the goalkeeper plays the same game when he faces a left-footed or a right-footed kicker, but the actions are simply identified differently. All that matters is whether the kicker and goalkeeper pick the kicker's strong side R or his weak side L. Payoffs are assumed to be the same for the two kicker types up to the renaming of the actions.[10]

Players in the sample choose either R or L 96·3% of the time, kickers 93·8% of the time and goalkeepers 98·9%.[11] For subtle "technological" reasons we will consider the choice C within their natural choices.[12] The typical penalty kick may then be described by the simple 2×2 model outlined in Section 2. Penalty kicks in this model have a unique Nash equilibrium and the equilibrium requires each player to use a mixed strategy. As mentioned earlier, equilibrium theory makes two testable predictions about the behaviour of kickers and goalkeepers: (1) winning probabilities should be the same across strategies for both players, and (2) each player's strategic choices must be serially independent.

Before we begin any formal test, it is worth examining the extent to which observed behaviour is close to the Nash equilibrium predictions. For all players in the sample the empirical scoring probabilities are

	g_L	$1 - g_L$
k_L	58·30	94·97
$1 - k_L$	92·91	69·92

where, as indicated above, k_L and g_L denote the non-natural sides. The mixed strategy Nash equilibrium predicted frequencies for these empirical values and the actual mixing probabilities observed in the sample are

	g_L (%)	$1 - g_L$ (%)	k_L (%)	$1 - k_L$ (%)
Nash predicted frequencies	41·99	58·01	38·54	61·46
Actual frequencies	42·31	57·69	39·98	60·02

10. As a referee noted, the assumption that the game is identical for the two kinds of kickers up to the renaming of the actions is not obvious. Hence, we have tested this assumption using a regression framework. The null hypothesis that kicker's types are perfectly symmetric corresponds to kicker-type fixed effects being jointly insignificant different from zero in different outcome variables: a goal is scored, the kicker shoots to the natural side, and the goalkeeper goes to the natural side. The analysis includes as explanatory variables several covariates that describe the state of the soccer match at the time the penalty is shot (match tied, goalkeepers' team ahead by one goal, ahead by two or more goals, and behind by one goal, goalkeeper is in the home team, time of the penalty), as well as goalkeeper-fixed effects. We find that in none of the cases can we reject the hypothesis that kicker types are identical. The p-values of the F-statistic for the joint significance of kicker-type fixed effects are 0·66, 0·63, and 0·72, respectively. Likewise, the hypothesis that goalkeepers facing different kicker types are identical cannot be rejected either. The p-values of the F-statistic for the joint significance of goalkeeper-fixed effects are 0·56, 0·61, and 0·59, respectively. Games are thus statistically identical across kicker types. The same conclusion can be reached using maximum likelihood techniques.

11. Chiappori *et al.* (2002) pay close attention to the possibility that C is an available pure strategy in their aggregate analysis and conclude that the availability of C as an action is not an issue. As indicated above (footnote 7), their findings are also substantiated in our data set. These results support the argument that a penalty kick may be described as a two-action game.

12. In personal interviews with professional players in the Spanish First Division, I was informed that they basically consider C and their natural strategy as equally natural. The reason is that they always kick with the interior side of their foot, which allows for greater control of the kick, by approaching the ball running from their non-natural side. This makes it equally difficult to shoot centre and to the natural side. See also Miller (1998) on this point, and on the starting position and the angle of run-up or approach to the ball by kicker type. There may be other alternative ways of treating the choice C. These are briefly discussed in Section 4.2. Not surprisingly, given the low frequency of this choice, these alternative ways have no effect on any of the results of the tests.

This evidence indicates that observed aggregate behaviour is virtually identical to the theoretical predictions.[13] We turn next to testing the implications of the Minimax theorem.

4.1. *Tests of equal scoring probabilities*

The tests of the null hypothesis that the scoring probabilities for both kickers and goalkeepers are identical across strategies can be implemented using Pearson's χ^2 goodness-of-fit test of equality of two distributions.

4.1.1. Individual tests. Let p^i_j denote the probability that player i will be successful when choosing strategy $j \in \{L, R\}$, n^i_j the number of times that i chooses j, and N^i_{jS} and N^i_{jF} the number of times in which kicker (goalkeeper) i is successful (S) or fails (F) at scoring (not scoring) the penalty kick, respectively, when choosing strategy j. Hence, under the null hypothesis $p^i_L = p^i_R = p^i$. When p^i is replaced by its maximum likelihood estimate $\frac{N^i_{LS}+N^i_{RS}}{n^i_L+n^i_R}$, then the Pearson statistic for player i

$$P^i = \sum\nolimits_{j\in\{L,R\}} \left[\frac{(N^i_{jS} - n^i_j p^i)^2}{n^i_j p^i} + \frac{(N^i_{jF} - n^i_j(1-p^i))^2}{n^i_j(1-p^i)} \right]$$

is distributed asymptotically as a χ^2 with 1 degree of freedom. The results of the tests are shown in Table 3.

The results show that the null hypothesis is not rejected for most players. Of the 42 players in the sample, the hypothesis is rejected for three players (two kickers and one goalkeeper) at the 5% significance level, and for five players (three kickers and two goalkeepers) at the 10% significance level. Note that with 42 players the expected number of rejections at the 5% level is 2·1 and at the 10% level is 4·2. These estimates suggest that at the individual level the hypothesis that scoring probabilities are identical across strategies cannot be rejected for most players at conventional significance levels. The number of rejections is basically identical, although slightly greater, than what the theory predicts, in particular for the subgroup of kickers.

Gary Lineker, a star English player in the 1980s and 1990s, describes penalty situations as "essentially a war of nerves between keeper and kicker" in Miller (1998). The descriptive evidence in Table 1 allows us to conjecture that nervousness may be a potential determinant of the scoring rate. This conjecture is intuitive and consistent with the fact that many superstars have missed or had penalties saved in critical situations of stress and pressure of great magnitude. It is also consistent with the evidence from other sports, even when players face no opponents and have no strategic choices to make (*e.g.* free throws in professional basketball).[14] However, even though nervousness may play a role in determining the scoring rate, it need not have any effect on the tests of equality of scoring probabilities. It would only play a role when the choice of strategy is related to the importance of the penalty. In order to evaluate the possible effects in the tests of equality of scoring probabilities, Palacios-Huerta (2002) considers the same subsample of

13. Observed behaviour is also remarkably close to the Nash predictions when penalties are sorted by kicker type (see Palacios-Huerta, 2002). This is not surprising given that games are statistically identical across types.

14. A logit regression for the scoring rate (a dichotomous indicator for the outcome of a penalty shot) confirms this conjecture for the penalties shot in the last 10 minutes of close games (when the score difference is 0, 1, or -1). Moreover, the data reveal that the scoring rate decreases at the end of a game not because goalkeeper's saves increase but because kickers shoot wide, to the goalpost, or to the crossbar more often than earlier in the game. This may be attributed to nervousness or to kickers being tired at the end of the game. There is no apparent way of determining which interpretation is the right one. However, nervousness is the typical interpretation offered by professional players in interviews (Miller, 1998), an interpretation that is consistent with the low level of effort that needs to be supplied in a penalty kick.

TABLE 3

Tests for equality of scoring probabilities

		Mixture		Scoring rates		Pearson	
Player	#Obs.	L	R	L	R	statistic	*p*-value
Kicker 1	34	0·32	0·68	0·91	0·91	0·000	0·970
Kicker 2	31	0·35	0·65	0·82	0·80	0·020	0·902
Kicker 3	40	0·48	0·52	0·74	0·76	0·030	0·855
Kicker 4	38	0·42	0·58	0·88	0·91	0·114	0·735
Kicker 5	38	0·50	0·50	0·79	0·84	0·175	0·676
Kicker 6	36	0·28	0·72	0·70	0·77	0·185	0·667
Kicker 7	41	0·20	0·80	0·75	0·82	0·191	0·662
Kicker 8	35	0·31	0·69	0·82	0·75	0·199	0·656
Kicker 9	31	0·19	0·81	0·83	0·92	0·416	0·519
Kicker 10	35	0·37	0·63	0·86	0·77	0·476	0·490
Kicker 11	32	0·48	0·52	0·87	0·94	0·521	0·471
Kicker 12	32	0·48	0·52	0·87	0·94	0·521	0·471
Kicker 13	38	0·55	0·45	0·76	0·88	0·907	0·341
Kicker 14	30	0·33	0·67	0·90	0·75	0·938	0·333
Kicker 15	30	0·50	0·50	0·80	0·93	1·154	0·283
Kicker 16	42	0·43	0·57	0·89	0·75	1·287	0·257
Kicker 17	40	0·42	0·58	0·58	0·85	1·637	0·201
Kicker 18	46	0·44	0·56	0·90	0·77	1·665	0·197
Kicker 19	39	0·48	0·52	0·74	0·90	1·761	0·184
Kicker 20	40	0·35	0·65	0·93	0·69	2·913	0·088*
Kicker 21	40	0·42	0·58	0·65	0·91	4·322	0·038**
Kicker 22	40	0·40	0·60	1·00	0·75	4·706	0·030**
All kickers	808	0·3998	0·6002	0·8111	0·8268		
Goalkeeper 1	37	0·38	0·62	0·21	0·22	0·000	0·982
Goalkeeper 2	38	0·39	0·61	0·20	0·22	0·017	0·898
Goalkeeper 3	30	0·60	0·40	0·28	0·25	0·028	0·866
Goalkeeper 4	50	0·46	0·54	0·17	0·15	0·061	0·804
Goalkeeper 5	36	0·33	0·67	0·25	0·21	0·080	0·777
Goalkeeper 6	34	0·44	0·56	0·27	0·21	0·147	0·702
Goalkeeper 7	37	0·19	0·81	0·14	0·10	0·221	0·638
Goalkeeper 8	37	0·54	0·46	0·25	0·18	0·293	0·588
Goalkeeper 9	32	0·56	0·44	0·22	0·14	0·326	0·568
Goalkeeper 10	40	0·45	0·55	0·11	0·18	0·388	0·533
Goalkeeper 11	33	0·18	0·82	0·17	0·30	0·416	0·519
Goalkeeper 12	30	0·27	0·73	0·25	0·14	0·545	0·460
Goalkeeper 13	34	0·41	0·59	0·14	0·25	0·578	0·447
Goalkeeper 14	40	0·50	0·50	0·15	0·25	0·625	0·429
Goalkeeper 15	44	0·45	0·55	0·10	0·21	0·957	0·328
Goalkeeper 16	36	0·31	0·69	0·09	0·24	1·804	0·298
Goalkeeper 17	42	0·55	0·45	0·30	0·11	2·449	0·118
Goalkeeper 18	42	0·38	0·62	0·13	0·35	2·506	0·113
Goalkeeper 19	42	0·40	0·60	0·35	0·12	3·261	0·071*
Goalkeeper 20	40	0·60	0·40	0·08	0·37	5·104	0·024**
All goalkeepers	754	0·4231	0·5769	0·1943	0·2068		

Note: *Indicates rejected at 10% level, and **indicates rejected at 5% level.

kickers and goalkeepers *except* "important penalties" (those penalties shot in the last 10 minutes of a match when the score difference was 0, 1, or −1). These penalties represent 12·9% of the sample. We then perform the same statistical tests at individual levels as in Table 3. The results show that the number of individual rejections decreases slightly: the null hypothesis is rejected for two players at the 5% level and for four players at the 10% level, again virtually identical to the 2·1 and 4·2 expected number of rejections in the sample predicted by the theory.

TABLE 4

Tests for equality of scoring probabilities for aggregate distributions

Panel A: Pearson tests

Tests of the joint hypothesis that the data for all experiments were generated by equilibrium play: $p^i_L = p^i_R$ for each player i.

	Pearson statistic	Degrees of freedom	p-value
All players	43·944	42	0·389
All kickers	24·138	22	0·340
All goalkeepers	19·806	20	0·470

Panel B: KS tests

Tests the null hypothesis that the empirical distribution of p-values in individual Pearson tests was generated by random draws from the uniform distribution $U[0,1]$.

	KS statistic	p-value
All players	0·527	0·883
All kickers	0·396	0·891
All goalkeepers	0·373	0·832

4.1.2. Aggregate tests. We next examine whether behaviour at the aggregate level can be considered to be generated from equilibrium play 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 in this case is the sum of the individual test statistics P^i. Under the null hypothesis this test is distributed as a χ^2 with 42 degrees of freedom. Note that this joint test allows for differences in probabilities p^i across players. The results are shown in Table 4.

Panel A shows that the Pearson statistic is 43·944 and its associated p-value is 0·389. This indicates that the null hypothesis that the data for all players were generated by equilibrium play cannot be rejected at conventional significance levels. If kickers and goalkeepers are considered as separate groups then under the null hypothesis the test is distributed as a χ^2 with 22 and 20 degrees of freedom respectively. For kickers the Pearson statistic is 24·138 and its p-value is 0·340, whereas for goalkeepers the Pearson statistic is 19·806 with a p-value of 0·470. The hypothesis of equality of winning probabilities also cannot be rejected for either subgroup.

As is well known, however, a potential problem with the Pearson joint test is that it may have little power against alternative hypotheses about how the data were generated. Panel B in this table also includes a more powerful test of the extent to which aggregate behaviour is consistent with the theory. As Walker and Wooders (2001) note, under the joint null hypothesis that all observations were generated by equilibrium play the p-values associated with the realized P^i statistics should be 42 draws from the uniform distribution $U[0, 1]$. A simple visual comparison of the distribution of p-values in Table 3 would initially appear to suggest that the data may be consistent with the theory since p-values are distributed quite uniformly across deciles. A formal assessment can be made by comparing the distribution of p-values with the uniform distribution using the Kolmogorov–Smirnov (KS) test. The value of the KS statistic for all players considered simultaneously is 0·527, with a p-value of 0·883. When kickers alone are considered simultaneously the KS statistic is 0·396, with a p-value of 0·891; for goalkeepers alone the KS statistic is 0·373 with an associated p-value of 0·832.

The results of these tests indicate more decisively than the Pearson's joint test that Minimax play has generated the data in the sample. In all cases the p-values do not even come close to rejecting the hypothesis of equality of winning probabilities across strategies.[15] In addition, as will be shown in the next section, the tests have substantial statistical power to distinguish equilibrium mixed strategy play from disequilibrium strategies. In consequence, we take the results in Tables 3 and 4 to be consistent with the first hypothesis of Minimax play.

4.1.3. Interpretation and discussion. As in all strictly competitive games, players' preferences are diametrically opposed in a penalty kick. The above results suggest that the actions taken by each penalty taker and each goalkeeper in the sample may be interpreted as a *maxminimizer* for each player in the sense that their actions maximize the payoffs that they can guarantee. As is well known, a player maxminimizes if he chooses an action that is best for him on the assumption that contingent upon his action his opponent will choose an action to hurt him as much as possible. The empirical evidence on professional penalty kicks is thus consistent with Nash equilibrium in this sense.

Notice that there are two special characteristics in this play that are not present in the typical experiment.

First, all penalties were not shot at the same time. In the data, most players are not involved in more than 15–20 penalties in a given season. Given that players and opponents may have limited information-processing ability and may be imperfect record keepers, players could deviate from Minimax play. However, the results of the tests suggest that they do not do so. The results show that they act instinctively and intuitively as if they were programmed with great preciseness to correctly play this strategic game. This may not be too surprising; after all, their knowledge and instincts have developed through years, often decades, by doing little else but playing soccer.

Second, opponents rotate. Under one interpretation given in the literature, the mixed strategy equilibrium provides a good description of the steady state behaviour of players who play *one* given game repeatedly against randomly selected opponents (Osborne and Rubinstein, 1994, pp. 38–39). Players, however, could feel freer to condition on their own past choices as an aid in achieving any desired move frequencies in these cases. However, the results show that they chose not to stray from Minimax play. A potential reason for their behaviour is that all players observe all choices of all opponents in the weekly television programmes and, as professionals, they do in fact keep written and mental records.[16] Moreover, as will be shown in the next subsection, tests of randomness show that their choices are serially independent; that is, they choose not to condition on past own or opponents' choices.

There are additional tests that can be implemented to confirm the idea that the mixed strategy equilibrium found in the data describes the steady state behaviour of a given player playing one given game against randomly selected opponents. For instance, consider Alan Shearer (player no. 2 in Table 3) from Newcastle United in the English Premier League. He is found to be indifferent between L and R strategies. Assume now that rather than having played one given game against his opponents he has actually played two different games, one against half of the opponents and

15. As indicated earlier, it is not possible to conduct individual level tests when few observations are available per player. In our sample the vast majority of players never choose C (particularly goalkeepers) or choose it just once or twice during the period 1995–2000. Also, as indicated above (footnote 11), the availability of C as a pure strategy is not an issue. Yet, the evidence is also consistent with Minimax play when individual Pearson tests and aggregate Pearson and KS tests are evaluated in a three-action game for kickers with at least three choices to the centre. The p-values of the individual tests are distributed quite uniformly across deciles, while the the p-value of the aggregate Pearson test is 0·820 and of the KS test is 0·501. These results are shown in Palacios-Huerta (2002).

16. Data on all professional players are relatively easy to obtain as virtually every penalty taken is televised. Miller (1998) and Anthony (2000) report interviews with professional players from the 1960s to the 1990s in England and Germany who acknowledged keeping written dossiers on their opponents, behaviour. This practice is now considered to be common among all professional players.

a second one against the other half. A simple test could then detect whether in fact more than one game was actually being played. Consider we take a subset of 30 out of the 41 observations we have for Alan Shearer. If in fact two different games were played—but somehow the tests of equality of winning rates in the whole sample did not reject the hypothesis of equality of winning rates across strategies—then testing the hypothesis for this subsample would tend to reject the hypothesis that winning rates are identical, or at least show significant variation in the p-values of the test. This idea is implemented in Palacios-Huerta (2002) for many different subsamples, all of the same size, randomly chosen for each player. The results show that the average p-values are extremely similar to those obtained in Table 3 for each and every kicker and goalkeeper in the sample. Moreover, the standard deviation of the p-values is also very low, always below 0·187. These results confirm the idea that for a given player one given game is being played against randomly selected opponents. This idea is also confirmed in a regression framework by testing for the homogeneity of the opponents for a given player, and by evaluating the stability of the payoff matrices for each and every player in the sample using the technique of bootstrapping (see Palacios-Huerta, 2002).

4.2. *Tests of serial independence*

The second testable implication is that a player's mixed strategy is the same at each penalty kick given constant payoffs across games. This implies that players' strategies are serially independent. More precisely, their play will not be serially independent if they choose not to switch their actions often enough or if they switch actions too often. In either case their play would not be consistent with the randomness implied by equilibrium play.

The work on randomization is now extensive in the experimental economics and psychological literatures. Interestingly, this hypothesis has *never* found support in any empirical (natural and experimental) tests of the Minimax hypothesis, and is rarely supported in other tests. 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 Boebel (1992), Rapoport and Budescu (1992), Mookherjee and Sopher (1994) and Camerer (1995)). Some subjects, however, have been taught to choose randomly after several hours of training in experimental settings (see Neuringer, 1986). These training data suggest that in some settings subjects might be able to learn to generate randomness. However, as Camerer (1995) remarks, "whether they do in other settings, under *natural* conditions, is an empirical question". We examine this open, elusive question next.

Consider the sequence of strategies chosen by player i in the order in which they occurred $s^i = \{s^i_1, s^i_2, \ldots, s^i_{n^i}\}$, where $s^i_x \in \{L, R\}$, $x \in [1, n^i]$, $n^i = n^i_L + n^i_R$, and n^i_R and n^i_L are the number of natural and non-natural choices made by player i. Let r^i denote the number of runs in the sequence s^i. A run is defined as a succession of one or more identical symbols which are followed and preceded by a different symbol or no symbol at all. Let $f(r^i; s^i)$ denote the probability that there are exactly r^i runs in the sequence s^i. Let $\Phi\left[r^i; s^i\right] = \sum_{k=1}^{r^i} f(k; s^i)$ denote the probability of obtaining r^i or fewer runs. Gibbons and Chakraborti (1992) show that by using the exact mean and variance of the number of runs in an ordered sequence then, under the null hypothesis that strategies are serially independent, the critical values for the rejection of the hypothesis can be found from the normal approximation to the null distribution. More precisely,

$$\frac{r^i - 0{\cdot}5 - 2(n^i_L n^i_R / n^i)}{\sqrt{2n^i_L n^i_R((2n^i_L n^i_R - n)(n^i)^2(n^i - 1))}}$$

TABLE 5

Tests of serial independence of choices

Player	Observations			Runs		
	L	R	Total	R	$\Phi[f(r-1;s)]$	$\Phi[f(r;s)]$
Kicker 1	11	23	34	16	0·439	0·597
Kicker 2	11	20	31	21	0·983**	0·994
Kicker 3	19	21	40	22	0·570	0·691
Kicker 4	16	22	38	19	0·365	0·496
Kicker 5	19	19	38	22	0·689	0·795
Kicker 6	10	26	36	15	0·344	0·509
Kicker 7	8	33	41	14	0·423	0·625
Kicker 8	11	24	35	15	0·263	0·407
Kicker 9	6	25	31	9	0·097	0·241
Kicker 10	13	22	35	19	0·599	0·729
Kicker 11	15	17	32	19	0·714	0·822
Kicker 12	15	17	32	20	0·822	0·901
Kicker 13	21	17	38	23	0·816	0·891
Kicker 14	10	20	30	12	0·117	0·221
Kicker 15	15	15	30	18	0·711	0·824
Kicker 16	18	24	42	19	0·164	0·254
Kicker 17	19	21	40	20	0·321	0·443
Kicker 18	20	26	46	19	0·693	0·789
Kicker 19	19	20	39	19	0·259	0·374
Kicker 20	14	26	40	14	0·022	0·049*
Kicker 21	17	23	40	18	0·159	0·251
Kicker 22	16	24	40	22	0·668	0·779
Goalkeeper 1	14	23	37	17	0·249	0·374
Goalkeeper 2	15	23	38	21	0·678	0·790
Goalkeeper 3	18	12	30	12	0·065	0·130
Goalkeeper 4	23	27	50	24	0·250	0·350
Goalkeeper 5	12	24	36	17	0·424	0·576
Goalkeeper 6	15	19	34	15	0·124	0·212
Goalkeeper 7	7	30	37	13	0·533	0·738
Goalkeeper 8	20	17	37	20	0·516	0·647
Goalkeeper 9	18	14	32	19	0·739	0·842
Goalkeeper 10	18	22	40	14	0·009	0·021**
Goalkeeper 11	6	27	33	11	0·423	0·661
Goalkeeper 12	8	22	30	15	0·802	0·908
Goalkeeper 13	14	20	34	19	0·644	0·767
Goalkeeper 14	20	20	40	22	0·564	0·685
Goalkeeper 15	20	24	44	27	0·871	0·925
Goalkeeper 16	11	25	36	16	0·378	0·535
Goalkeeper 17	23	19	42	28	0·964*	0·983
Goalkeeper 18	16	26	42	23	0·713	0·814
Goalkeeper 19	17	25	42	18	0·113	0·187
Goalkeeper 20	24	16	40	19	0·285	0·408

Note: *Indicates rejected at 10% level, and **indicates rejected at 5% level.

is distributed as a standardized normal probability distribution. The null hypothesis will then be rejected at the 5% 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 $\Phi[r^i; s^i] < 0{\cdot}025$ or if $1 - \Phi[r^i - 1; s^i] < 0{\cdot}025$. The results of these tests of serial independence are shown in Table 5.

These results show that the null hypothesis of serial independence is rejected for two players (one kicker and one goalkeeper) at the 5% significance level. With 42 players the expected number of rejections at this level is 2·1. At the 10% significance level, the number of rejections is four (two kickers and two goalkeepers) with 4·2 being the expected number of rejections.

TABLE 6

Results of significance tests from logit equations for the choice of the natural side

Estimating equation:

$$R = G[a_0 + a_1\text{lag}(R) + a_2\text{lag2}(R) + b_0 R^* + b_1\text{lag}(R^*) + b_2\text{lag2}(R^*) + c_1\text{lag}(R)\text{lag}(R^*) + c_2\text{lag2}(R)\text{lag2}(R^*)]$$

		Players whose behaviour allows rejection of the null hypothesis at the:		
Null hypothesis:		0·05 level	0·10 level	0·20 level
1. $a_1 = a_2 = b_0 = b_1 = b_2 = c_1 = c_2 = 0$	Kicker	—	2	2, 18
	Goalkeeper	—	7	7, 15
2. $a_1 = a_2 = 0$	Kicker	—	2	2, 14
	Goalkeeper	—	8	8, 17
3. $b_1 = b_2 = 0$	Kicker	—	—	5
	Goalkeeper	—	7	7
4. $c_1 = c_2 = 0$	Kicker	—	—	6
	Goalkeeper	—	—	14
5. $b_0 = 0$	Kicker	—	11, 17	5, 11, 17, 21
	Goalkeeper	—	3, 16	3, 9, 10, 16

Notes: R and R^* denote the choice of "natural" strategy by a kicker and a goalkeeper, respectively (right for a right-footed kicker and for a goalkeeper facing a right-footed kicker, and left for a left-footed kicker and for a goalkeeper facing a left-footed kicker). The terms "lag" and "lag2" refer to the strategies previously followed in the ordered sequence of penalty kicks. $G[x]$ denotes the function $\exp(x)/[1 + \exp(x)]$. Rejections are based on likelihood-ratio tests.

These findings suggest that professional soccer players are indeed able to generate random sequences; they neither switch strategies too often nor too little. The number of rejections is remarkably consistent with the theory. As indicated above, this result is in sharp contrast with the overwhelming experimental evidence from the psychological and experimental literatures mentioned earlier, and also with the evidence from first serves in tennis players (Walker and Wooders, 2001).

Note also that the values in columns $\Phi\left[r^i; s^i\right]$ and $\Phi\left[r^i - 1; s^i\right]$ tend to be uniformly distributed in the [0, 1] interval. This suggests that professional soccer players do not even have a (statistically insignificant) tendency to "sit on" a given strategy or to switch strategies too often. To confirm that past choices have no role in determining current choices, we follow the analysis in Brown and Rosenthal (1990) and estimate a logit equation for each player. The dependent variable is a dichotomous indicator of the choice of natural side. The independent variables are first and second lagged indicators for both players' past choices, first and second lags for the product of their choices, and an indicator for the opponent's current choices. The results are shown in Table 6.

The main result in this table is that the null hypothesis that *all* the explanatory variables are jointly statistically insignificant (hypothesis #1) cannot be rejected for any player at the 5% level, and is rejected for only two players at the 10% level.

The table also reports the tests of different hypotheses concerning whether one's past choices alone, past opponent's choices alone, and successful past plays alone may determine current choices (hypotheses #2 through #4). No evidence that any player made choices in a serially dependent fashion in any respect is found at the 5% level, while at the 10% level none of the hypotheses are rejected for more than two players. As suggested by the findings in Table 5, these results indicate that the choices of most players are unrelated to their own previous choices and outcomes, and to opponents' previous choices and outcomes.

An interesting finding is concerned with the role of the opponent's *current* choice (hypothesis #5). As argued in Section 2, the 12-yard distance between the penalty mark and the goal line is too short for players to choose not to move simultaneously. Consistent with this idea, the coefficient b_0 in Table 6 that captures the role of the opponent's current choice is not significant for any player at the 5% level, and for four players at the 10% level. At the 20% level it is significant for about four kickers and four goalkeepers. Although this confidence level is somewhat generous by conventional standards, these results could suggest that there may be a small element of sequentiality in some players' choices or, perhaps, some ability "to read the opponent's face".[17]

Finally, we may also test the joint hypothesis that each of the 42 experiments is serially independent. Following the suggestion in Walker and Wooders (2001), we employ the KS goodness-of-fit test by constructing a random draw d^i from the uniform distribution $U\left[\Phi\left[r^i-1;s^i\right],\Phi\left[r^i;s^i\right]\right]$ for each player i.[18] Under the null hypothesis of serial independence d^i is distributed as a $U[0,1]$. Figure 1 shows how the empirical cumulative distribution function of a particular realization is strikingly similar to the theoretical prediction. After performing 10,000 trials with such random draws for each player, the average value of the KS test statistic that compares the cumulative distribution of the realized values d^i with the uniform distribution is 0·660 with a standard deviation of 0·005. The average p-value is 0·780 with a standard deviation of 0·001. Hence, the hypothesis that each of the 42 experiments is serially independent cannot be rejected. Similar results are also obtained for the subgroups of kickers and goalkeepers.

We take the results of the tests of randomness as consistent with the hypothesis that the strategies followed by professional soccer players are serially independent. This evidence represents the first time that individuals have been found to display statistically significant serial independence in a strategic game in a natural setting.

5. DISCUSSION AND ADDITIONAL EVIDENCE

Empirical evidence from the behaviour of professional soccer players in penalty kicks provides substantial support for the two empirical implications that derive from the hypothesis that agents play according to equilibrium. Clearly, the positive results obtained from the analysis may in large part be attributed to the distinct virtues of the natural play. This one-shot, face-to-face play involves professional subjects. These subjects have had the time necessary to become proficient at generating random sequences and to develop the instincts and learn what is considered to be the correct way of playing two-person zero-sum games. Motivation and incentives are high, and strategic choices and all relevant features of the environment are observed. The expertise of the players and the simple features of the environment that approximate so well the setting in the theoretical model may also explain why positive results for the Minimax theorem are obtained in this natural setting with a smaller sample size than the amount of data usually elicited from subjects in experimental settings. Hence, professional soccer players strictly focus on maximizing their payoffs in penalty kicks and are not involved in exploring non-equilibrium alternative patterns of behaviour in their play.

The analysis also provides various advantages over previous attempts to test for Minimax play and mixed strategies in natural conditions in the literature. As indicated earlier, Chiappori *et al*. (2002) are concerned with problems of aggregation when few observations are available per player and the Minimax equilibrium implications for *individual* players, including serial independence, cannot be tested. They show that it is still possible to test for various predictions

17. A small element of sequentiality may induce players to choose C. For instance, kickers may choose C if the goalkeeper moves too early, and goalkeepers may choose it if they react too late. None of the results of the tests of equality of winning rates and randomness change in a noticeable way if the observations C for kickers are interpreted this way (as the choice opposite to the one taken by the goalkeeper). For goalkeepers the number of C observations is basically negligible for different interpretations to induce differences in the tests.

18. This procedure is able to circumvent the fact that the KS cannot be applied directly to the values in the columns $\Phi(r^i;s^i)$ and $\Phi(r^i-1;s^i)$ given that their values are neither identically distributed nor continuously distributed.

that arise when aggregating observations across players. As indicated earlier the basic aggregate predictions of their model are also substantiated in our data set. With respect to Walker and Wooders (2001), an important feature of the analysis in this paper is the ability to test equilibrium restrictions on *both* sides of the game. Win rates for both players in the game can be computed, which turn out to be statistically identical across strategies. Walker and Wooders (2001) can compute win rates for only one of the two players (the server). It is thus conceivable that win rates for players receiving serves are different and, hence, that servers are not mixing properly. This aspect cannot be tested in their data. A second relevant feature is that by using *all* penalty kicks we avoid a subtle potential problem of selection bias that may be present in Walker and Wooders (2001). The authors choose tennis matches that are long enough to provide lots of observations of L and R services. However, by choosing only long matches they may be selecting for matches in which players are mixing optimally and, hence, overstating the degree of conformity with the first implication of the Minimax theorem. By using all penalty kicks, the analysis in this paper does not have such a potential problem of selection bias. A third aspect is that penalty kicks are a simultaneous-move game since outcomes are decided immediately. Hence, there is no subsequent strategic play that would need to be modelled theoretically and addressed empirically as in tennis or in other non-simultaneous games. Finally, strategic choices exhibit serial independence and hence conform with the second implication of the Minimax theorem. Although this result may appear intuitive in our natural setting, it has not been obtained previously under natural or experimental conditions.

In the remainder of this section we show and discuss other evidence that complements the empirical findings. We conduct Monte Carlo simulations to evaluate the power of the Pearson and Kolmogorov–Smirnov tests of equality of winning probabilities to reject the null hypothesis when various alternative hypotheses are true. We also discuss related evidence on the result of serial independence.

5.1. *Power of the tests*

The data were found to support the Minimax theorem but, in principle, they may also support other models having little to do with Minimax play. We examine the ability of the tests to reject the null hypothesis when various alternative hypotheses are true by evaluating the power of the Pearson and KS tests of equality of winning probabilities to detect deviations from Minimax play using Monte Carlo simulations.

Recall that the actual mixing behaviour, $k_L = 39{\cdot}98\%$ and $1-k_L = 60{\cdot}02\%$ for kickers, and $g_L = 42{\cdot}31\%$, $1 - g_L = 57{\cdot}69\%$ for goalkeepers, is basically identical to the Nash equilibrium predicted frequencies: $k_L = 38{\cdot}54\%$, $1 - k_L = 61{\cdot}46\%$, $g_L = 41{\cdot}99\%$, $1 - g_L = 58{\cdot}01\%$. This evidence may initially suggest that the tests are going to have substantial power.

Under the null hypothesis that in each case opponents follow their equilibrium mixture, the Pearson test statistic $P^n = \sum_{i=1}^{n} P^i$ is distributed as a χ^2 with n degrees of freedom ($n = 22$ for kickers and $n = 20$ for goalkeepers). We compute the power functions for these tests by randomly generating 100,000 times the data for each of the "experiments" in the initial subsample—assuming the average frequencies with which the natural and non-natural sides are chosen—and computing the frequency with which the Pearson joint test rejects the null hypothesis against a true alternative hypothesis at the 5% significance level. The power functions are depicted in Figure 2.

The results of the simulations show that the tests have substantial power for both kickers and goalkeepers. For instance, if opponents were assumed to choose their strategies with equal probability the null hypothesis would be rejected with probability 0·70 for kickers and 0·92 for goalkeepers; if they were assumed to choose the natural side 70% of the time the null hypothesis

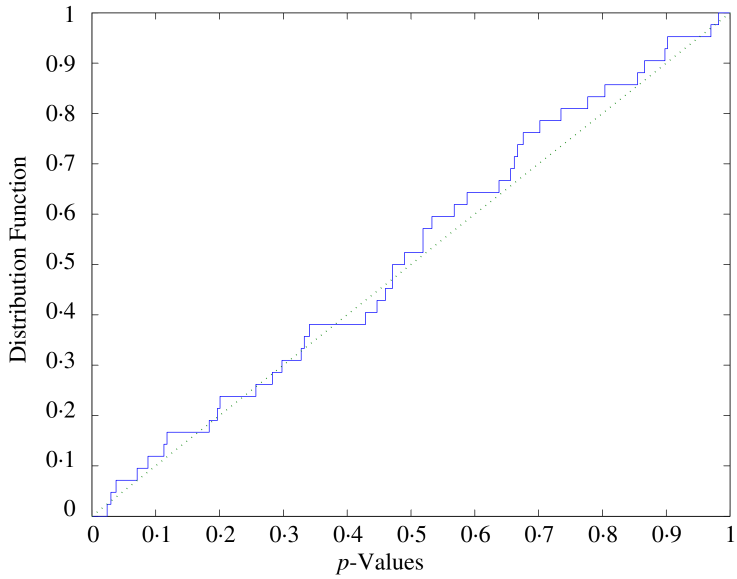

FIGURE 1

KS test of serial independence

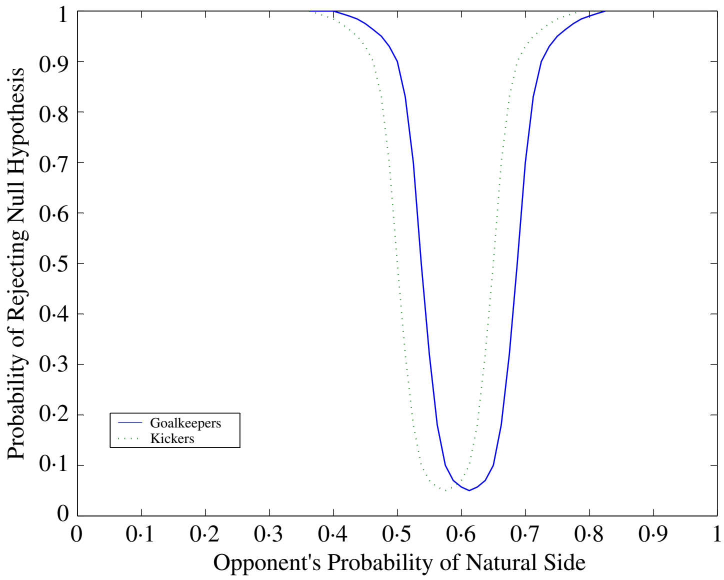

FIGURE 2

Power functions

would be rejected with probability 0·80 for goalkeepers and 0·95 for kickers.[19] Likewise, the KS joint tests of equality of winning probabilities also show significant power to reject the

19. The power functions have also been computed for left-footed kickers alone, right-footed kickers alone, and for the goalkeepers facing these types of kickers respectively. The corresponding functions are very similar to those presented in Figure 2.

null hypothesis when alternative hypotheses are true. The power functions are in fact basically identical.

5.2. *Serial independence and time delays*

As is well known, the testable implications of the Minimax theorem examined in this paper and in the literature are unaffected by different time lags. More precisely, equilibrium strategies in repeated zero-sum games are independent of the time lags between the stages of the game. Moreover, the equilibrium strategies dictate that at every stage players play according to the equilibrium strategy of the stage game. Therefore, from the theoretical perspective time lags have no influence on equilibrium play.

From the empirical perspective, however, it is at least conceivable that the lack of serial correlation in the sequences chosen by kickers and goalkeepers may have been caused, at least in part, by the actual time delay between penalties. Recall that in this natural scenario players are typically involved in no more that 15–20 penalties per season. Given that there is no evidence in the extensive literature on randomization about the effect of different time lags in zero-sum or other games, and that the hypothesis of serial independence has not been supported previously, it is of interest to study whether in this natural setting the time delay between penalties may have played a role in the serial independence result. Intuitively, however, the possible role is not clear. On one hand, time lags and changes in context may decay memory and thus help produce sequences that are random. On the other hand, as mentioned earlier (see footnote 16), professional players keep written records of opponents' choices in previous penalties; so, in a sense, they may have perfect memory, perhaps even better than if penalties were shot successively.

In order to study this aspect, Palacios-Huerta (2002) implements the tests of serial independence with data from penalty shootouts where kicks occur in rapid succession in short spans of time. The data come from elimination tournaments where ties are broken with penalty shootouts. The result of serial independence also appears for penalty shootouts. Hence, spacing strategies out over time or taking them in rapid succession does not appear to make any difference for these professional players to generate sequences with no serial correlation in this natural play.

We conclude that even though different time lags play no role in the theoretical implications of repeated zero-sum games, the empirical role they may play in empirical studies of strategic games in natural and experimental settings is an important, understudied aspect in the literature. It is thus an aspect that deserves careful analysis, especially given the fact that randomness is such an important but elusive phenomenon and that time lags may characterize many natural environments.

6. CONCLUDING REMARKS

Over the last few decades non-cooperative game theory has become a standard tool in economics and other social sciences. At the same time it has also come under increasing scrutiny from theoretical and experimental economists. Despite recent substantial progress in the literature, an important challenge facing non-cooperative game theory is that of providing compelling evidence that predictions are confirmed by empirical evidence, which in turn often means using the full extent of experimental and natural data to shape generalizations of game theory.

The analysis in this paper exploits a unique data set on a one-shot two-person zero-sum game involving expert players under natural conditions. The results of the tests are remarkably consistent with equilibrium play in every respect: (i) winning probabilities are statistically identical across strategies for players; (ii) players generate serially independent sequences and ignore any possible strategic links between plays. The tests have substantial power to distinguish equilibrium mixed strategy play from disequilibrium alternatives. These results represent, to the best of our knowledge, the first time that both implications of von Neumann's (1928) Minimax theorem are supported under natural conditions.

APPENDIX. IDENTITIES OF KICKERS AND GOALKEEPERS

Players are grouped by the country of the professional league where they played in at the end of the period of analysis. In brackets is the identification number used in Tables 3, 5, and 6, and in parenthesis is the professional team they played for.

KICKERS

Italy: [9] Batistuta (Roma), [13] Baggio (Brescia), [11] Del Piero (Juventus), [5] Mihajlovic (Lazio), [15] Chiesa (Fiorentina), [6] Signori (Bologna), [7] Rui Costa (AC Milan), [8] Amoroso (Udinese), [1] Mendieta (Lazio).

Spain: [22] Penev (Atletico de Madrid), [17] Hierro (Real Madrid), [16] Larrazabal (Athletic de Bilbao), [14] Garitano (Zaragoza), [19] Catanha (Celta), [20] Donosti (Eibar), [12] Juninho (Atletico de Madrid/Vasco de Gama), [10] Rivaldo (Barcelona), [3] Zidane (Real Madrid).

England: [2] Shearer (Newcastle), [4] Bergkamp (Arsenal), [21] Finidi (Ipswich Town), [18] Suker (West Ham).

GOALKEEPERS

Italy: [7] Toldo (Inter Milan), [9] Mazzantani (Perugia), [10] Peruzzi (Lazio), [14] Pagliuca (Bologna), [11] Taibi (Atalanta), [3] Brivio (Venezia), [12] Buffon (Juventus).

Spain: [2] Cesar (Real Madrid), [1] Alberto (Real Sociedad), [13] Cañizares (Valencia), [4] Ceballos (Racing de Santander), [17] Stelea (Salamanca), [18] Etxebarria (Rayo Vallecano), [8] Molina (Deportivo Coruña), [19] Juanmi (Zaragoza), [20] Dutruel (Barcelona), [16] Esteban (Oviedo), [5] Prats (Betis).

England: [15] Seaman (Arsenal), [6] Schmeichel (Aston Villa).

Notes: Kickers number 5, 6, 10, 14, 16, 18, and 22 are left-footed. All others are right-footed. Penev (Atletico de Madrid) retired in 2000.

Acknowledgements. I am grateful to Robert J. Aumann, Pierre-Andre Chiappori, James Dow (the editor), Peter Hansen, Tony Lancaster, Steve Levitt, Ana I. Saracho, Roberto Serrano and, especially, Oscar Volij and two anonymous referees for useful comments and suggestions. I am also grateful to seminar participants at Brown University, Iowa State University, the First World Congress of the Game Theory Society, and the V Spanish Meeting on Game Theory and Applications, to Subhalakashmi Ghosh and Maria Luz Huerta for able research assistance, and to the George J. Stigler Center and the Salomon Foundation for financial support. Oscar Volij suggested the title. Any errors are mine alone.

REFERENCES

ANTHONY, A. (2000) *On Penalties* (London: Yellow Jersey Press).

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

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

CAMERER, C. (1995), "Individual Decision Making", in John H. Hagel and Alvin E. Roth (eds.) *Handbook of Experimental Economics* (Princeton, NJ: Princeton University Press) 587–703.

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

CAMERER, C. and HO, T.-H. (1999), "Experience-Weighted Attraction Learning in Normal Form Games", *Econometrica*, **67**, 827–874.

CHIAPPORI, P. A., LEVITT, S. D. and GROSECLOSE, T. (2002), "Testing Mixed Strategy Equilibrium When Players are Heterogeneous: The Case of PenaltyKicks in Soccer", *American Economic Review*, **92**, 1138–1151.

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

FÉDÉRATION INTERNATIONALE DE FOOTBALL ASSOCIATION (2000) *Official Laws of the Game* (Chicago, IL: Triumph Books).

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

KREPS, D. M. (1991) *Game Theory and Economic Modelling* (Oxford: Oxford University Press).

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

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

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

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

OSBORNE, M. J. and RUBINSTEIN, A. (1994) *A Course in Game Theory* (Cambridge, MA: MIT Press).

PALACIOS-HUERTA, I. (2002), "Professionals Play Minimax: Appendix" (Mimeo, Brown University).

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

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

STAHL, D. O. (2000), "Rule Learning in Symmetric Normal-Form Games: Theory and Evidence", *Games and Economic Behavior*, **32**, 105–138.

VON NEUMANN, J. (1928), "Zur Theorie der Gesellschaftsspiele", *Mathematische Annalen*, **100**, 295–320.

WALKER, M. and WOODERS, J. (2001), "Minimax Play at Wimbledon", *American Economic Review*, **91**, 1521–1538.