STRICTLY COMPETITIVE STRATEGIC SITUATIONS

IGNACIO PALACIOS-HUERTA (1)
London School of Economics
and Ikerbasque Foundation at UPV/EHU

«I thought there was nothing worth publishing until the Minimax Theorem was proved. As far as I can see, there could be no theory of games without that theorem.»
JOHN VON NEUMANN, 1953

«Strictly competitive strategic situations are a vital cornerstone of game theory.»
ROBERT J. AUMANN, 1987

Consider oligopoly firms' dynamic pricing strategies in a gasoline market under a law that constrains firms to set price simultaneously and only once per day (Wang, 2009). Consider firms interested in monitoring whether or not their employees are working, and the typical employee decision whether to work or shirk at a given time. Consider the tax authority's decision whether or not to audit a tax payer, and the tax payer's decision whether or not to cheat on his taxes. Consider a professional cyclist in the Tour de France deciding whether to dope or not. Consider the problem of parking legally or illegally in cities, and the police deciding whether to monitor or not a given street at a point in time. Consider a penalty kick in football, where both kicker and goalkeeper are deciding simultaneously which side to take. These are all competitive situations that involve «mixed strategies», that is situations in which optimal behaviour involves mixing among the various alternatives available to agents (firms, employees, drivers, police, tax authorities).

This paper reviews the economic literature dealing with the experimental testing of situations where two agents compete strategically in a strict sense of the word, that is in a zero-sum fashion in which the gain to one is exactly identical to the loss of the other. These situations are important for economics and the social sciences because they are at the intersection of two of the most fundamental concepts in Economics: «competition» and «strategic behaviour» in the sense of interactive decision-making. The Merriam-Webster dictionary defines the first concept as follows:

Compete, competition: to strive consciously or unconsciously for an objective (as position, profit, or a prize); to be in a state of rivalry «*competing* teams, companies *competing* for customers».

And Rubinstein (1991), defines the concept of «equilibrium strategy» as follows:

Equilibrium Strategy. It describes an agent plan of action when dealing with situations of conflict as well as those considerations which support the optimality of his plan.

HISTORY AND RELEVANCE IN ECONOMICS

Kreps (1991) correctly notes that «the point of game theory is to help economists understand and predict what will happen in economic, social and political contexts». So if Von Neumann considered, as the initial quotation suggests, that there could be no theory of games without proving the Minimax theorem, then it seems appropriate to think that he would have considered that there could be no empirical applicability of the theory of games without first having verified empirically that theorem. As will be noted below, the Minimax theorem was not empirically verified until 2003, that is 75 years after first formally demonstrated in 1928.

The empirical verification of strategic models of behavior is often difficult and problematic. In fact, testing the implications of *any* game theoretical model in a real-life setting has proven extremely difficult in the economics literature for a number of reasons. The primary reason is that many predictions often hinge on properties of the utility functions and the values of the rewards used. Further, 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 the actual strategies available to the individuals or firms involved, as well as in measuring choices, effort levels, and the incentive structures they face. As a result, even the most fundamental predictions of game-theoretical models have rarely been supported empirically in real situations.

Von Neumann published the Minimax theorem in his 1928 article «Zur Theorie der Gesellschaftsspiele». The theorem essentially says:

For every two-agent, zero-sum game with finitely many strategies, there exists a value V and a mixed strategy for each agent or player, such that:

(a) Given player 2's strategy, the best payoff possible for player 1 is V, and

(b) Given player 1's strategy, the best payoff possible for player 2 is −V.

Equivalently, Player 1's strategy guarantees himself a payoff of V regardless of Player 2's strategy, and similarly Player 2 can guarantee himself a payoff of −V.

A mixed strategy is a strategy consisting of possible moves and a probability distribution (collection of weights) that corresponds to how frequently each move or potential alternative is to be played. Interestingly, there are a number of interpretations of mixed strategy equilibrium, and economists often disagree as to which one is the most appropriate. See for example the interesting discussion in the classic graduate textbook by Martin Osborne and Ariel Rubinstein, *A Course on Game Theory* (1994, Section 3.2).

A game is called zero-sum or, more generally, constant-sum, if the two players' payoffs always sum to a constant, the idea being that the payoff of one player is always exactly the same as the negative of that of the other player. The name Minimax arises because each player minimizes the maximum payoff possible for the other. Since the game is zero-sum, he also minimizes his own maximum loss (i.e., maximizes his minimum payoff).

Most games or strategic situations in reality involve a mixture of conflict and common interest. Sometimes, everyone wins, such as when players engage in voluntary trade for mutual benefit. In other situations, everyone can lose, as the well-known Prisoner's dilemma situations illustrate. Thus, the case of *pure conflict* (or zero-sum or constant-sum or strictly competitive) games represents the extreme case of conflict situations that involve no common interest. As such, and as Aumann (1987) puts it in the initial quote, zero-sum games are a key cornerstone of game theory. It is not a surprise that they were the first to be studied theoretically.

The Minimax theorem can be regarded as a special case of the more general theory of Nash (1950, 1951). It applies only to two-person zero-sum or constant-sum games, while the Nash equilibrium concept can be used with any number of players and any mixture of conflict and common interest in the game.

A final characteristic of the mixed strategy equilibrium in zero‐sum games is that it is relatively intuitive since players in such games have the incentive to deliberately randomize to remain unpredictable. However, the use of mixed strategies in nonzero‐sum games is rather counterintuitive since players in such games often have the incentive to avoid randomization (Schelling 1960).

Rubinstein (1991) notes that there is an «enormous economic literature that utilizes mixed strategy equilibrium».

FIRST EMPIRICAL VERIFICATION OF THE MINIMAX THEOREM

In what follows I will consider a simple penalty kick situation in football both for the sake of expositional clarity, and because, as will be seen, this setting turned out to the one in which the first complete empirical verification of the Minimax theorem in real life was obtained.

A formal model of the penalty kick may be written as follows. 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. Take, 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:

TABLE 1

i\j	L	R
L	π_{LL}, $1-\pi_{LL}$	π_{LR}, $1-\pi_{LR}$
R	π_{RL}, $1-\pi_{RL}$	π_{RR}, $1-\pi_{RR}$

Each penalty kick involves two players: a kicker and a goalkeeper. In the typical kick in professional leagues the ball takes about 0.3 seconds to travel the distance between the penalty mark and the goal line. This is less time than it takes for the goalkeeper to react and move to the possible paths of the ball. Hence, both kicker and goalkeeper must move simultaneously. Players have few strategies available and their actions are observable. There are no second penalties in the event that a goal is not scored. The initial location of both the ball and the goalkeeper is always the same: the ball is placed on the penalty mark and the goalkeeper positions himself on the goal line, equidistant from the goalposts. The outcome is decided, in effect, immediately (roughly within 0.3 seconds) after players choose their strategies.

The clarity of the rules and the detailed structure of this simultaneous one-shot play capture the theoretical setting of a zero-sum game extremely well. In this sense, and as will be discussed below, it presents notable advantages over other plays in professional sports and other real-world settings.

The penalty kick game has a unique Nash equilibrium in mixed strategies when:

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

Testable Implications. If the play in a penalty kick can be represented by this model, then equilibrium play requires each player to use a mixed strategy. In this case, the equilibrium yields two sharp testable predictions about the behavior 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 probability of success identical across strategies. That is, g_L should satisfy $pk_L = pk_R$, where:

$$pk_L = g_L \cdot \pi_{LL} + (1 - g_L) \cdot \pi_{LR}$$
$$pk_R = g_L \cdot \pi_{RL} + (1 - g_L) \cdot \pi_{RR}$$

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, $pg_L = pg_R$, where:

$$pg_L = k_L \cdot (1 - \pi_{LL}) + (1 - k_L) \cdot (1 - \pi_{RL})$$
$$pg_R = k_L \cdot (1 - \pi_{LR}) + (1 - k_L) \cdot (1 - \pi_{RR})$$

2. Each player's choices must be serially independent given constant payoffs across games (penalty kicks). That is, individuals must be concerned only with instantaneous payoffs and intertemporal links between penalty kicks 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.

The intuition for these two testable hypotheses is the following. In a game of pure conflict (zero-sum), if it would be disadvantageous for you to let your opponent see your actual choice in advance, then you benefit by choosing at random from your available pure strategies. The proportions in your mix should be such that the opponent cannot exploit your choice by pursuing any particular pure strategy out of those available to him—that is, each player should get the same average payoff when he plays any of his pure strategies against his opponent's mixture.

Note that theories that make precise point predictions are easy to reject. In our case, agents have to choose the *exact* frequencies of actions, to *exactly* equate payoffs across strategies, and they have to do that in an *exactly* random way. Even minimal deviation from the frequencies and payoffs, or from randomization are sufficient for rejecting the theory. So this reject/no reject dichotomy is quite rigid. O'Neill (1991) suggests for these cases an alternative that is much less rigid than the reject/no reject dichotomy: a Bayesian approach to hypothesis testing combined with a measure of closeness of the results to the predictions.

Advantages over other Empirical Settings. Although sports settings have many advantages, it is often impossible to find settings with the same ideal characteristics of a penalty kick. In fact even seemingly similar situations often *deviate substantially* from the theoretical postulates because either they are not always the same game when repeated and/or they are not even a simultaneous zero-sum game:

– *Not always the same game*. Take serving in tennis or pitching in baseball. Not only the direction of the serve or the pitch, but its spin are also important choice variables that cannot be dismissed. More importantly, the position of the player returning a tennis serve or attempting to hit a baseball ball affects the choice of strategy by the server or the pitcher. Thus, unlike what happens in a penalty kick or in a lab experiment, the initial location of the two players makes the basic situation vary from game to game. As a result games are different and cannot be aggregated to compute average frequencies or study randomness.

– *Not a simultaneous zero-sum game*. Perhaps the most critical shortcoming is that a serve or a pitch is not a simultaneous (static) but a sequential (dynamic) game, in that the outcome of the play is typically not decided immediately. After a tennis player serves or a pitcher throws, there is subsequent strategic play that often plays a crucial role in determining the final outcome. Each point in these situations is more like part of a dynamic game with learning, where each player plays a multi-armed bandit at the start of the match. In a dynamic game, there probably are spillovers from point to point whereas in a standard repeated zero-sum game, especially if repeated infrequently, there are no such payoff spillovers. For instance, in tennis having served to the left on the first serve (and say faulted) may effectively be «practice» in a way that makes the server momentarily better than the average at serving to the left again. If this effect is important, the probability that the next serve should be down the line should increase. In other words, there should be negative serial correlation in the choice of serve strategies rather than the lack of serial correlation predicted by Minimax.

Consistent with these shortcomings, the results in Walker and Wooders (2001) confirm that tennis players switch serving strategies too often to be consistent with random play, and hence with Minimax. This is also confirmed in Schweizer (2016) and Spiliopoulos (2016) who show that tennis players do not even equate payoffs across strategies.

TESTING

A penalty kick that is described by the simple 2×2 model (2) has a *unique* Nash equilibrium and the equilibrium requires each player to use a mixed strategy. As just noted above, equilibrium theory makes two testable predictions about the behavior 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.

Using data collected on 9,017 penalty kicks during the period September 1995-June 2012 from professional games in Spain, Italy, England, and other countries, the empirical scoring and no-scoring probabilities in percentage terms are:

$i \backslash jC$	g_L	$1-g_L$
k_L	59.11, 40.89	94.10, 5.90
$1-k_L$	93.10, 6.90	71.22, 28.78

If we compute the mixed strategy Nash equilibrium in this game (Minimax frequencies), and we look at the actual mixing probabilities observed in the sample, we find that observed aggregate behavior is virtually *identical* to the theoretical predictions:

	g_L	$1-g_L$	k_L	$1-k_L$
Nash Predicted Frequencies:	40.23%	59.77%	38.47%	61.53%
Actual Frequencies:	41.17%	58.83%	38.97%	61.03%

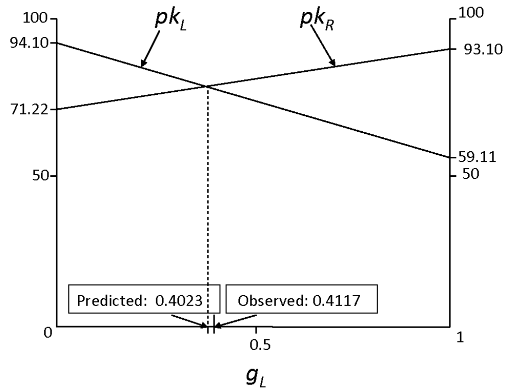

Figure 1 – Nash and Actual Frequencies for Goalkeepers

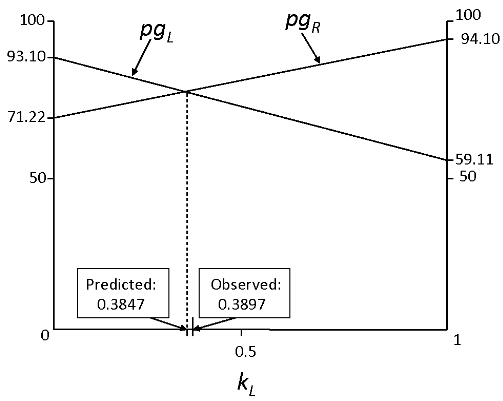

Figure 2 – Nash and Actual Frequencies for Kickers

This result is, at the very least, encouraging for the model. We turn next to testing the two implications of the Minimax Theorem.

Implication #1: *Tests of Equal Scoring Probabilities*

The tests of the null hypothesis that the scoring probabilities for a player (kicker or goalkeeper) are identical across strategies can be implemented with the standard proportions tests, that is, using Pearson's X^2 goodness-of-fit test of equality of two distributions.

Let p_{ij} denote the probability that player i will be successful when choosing strategy $j \in \{L,R\}$, n_{ij} the number of times that i chooses j, and N_{ijS} and N_{ijF} the number of times in which player i chooses strategy j and is successful (S) or fails (F) in the penalty kick. Success for a kicker is to score a goal, and for a goalkeeper is that a goal is not scored. Hence, we want to test the null hypothesis $p_{iL} = p_{iR} = p_i$. Statisticians tell us that to do this, the Pearson statistic for player i

$$P^i = \sum_{j\in\{L,R\}} \left[\frac{\left(N_{ijS} - n_{ij}p_i\right)^2}{n_{ij}p_i} + \frac{\left(N_{ijF} - n_{ij}\left(1-p_i\right)\right)^2}{n_{ij}(1-p_i)} \right]$$

is distributed asymptotically as a X^2 with 1 degree of freedom.

It is also possible to study whether behavior at the aggregate level is consistent with equilibrium play by testing the joint hypothesis that each individual case is *simultaneously* generated by equilibrium play. The test statistic for the Pearson joint test in this case is the sum of all the N individual test statistics, and under the null hypothesis this test is distributed as a X^2 with N degrees of freedom. Note that this joint test allows for differences in probabilities P^i across players.

Implication #2: *Tests of Randomness or Serial Independence*

The second testable implication is that a player's mixed strategy is the same at each penalty kick. This implies that players' strategies are random or serially independent. Their play will not be serially independent if, for instance, they choose not to switch their actions often enough or if they switch actions too often or if they follow any other non-random pattern.

The work on randomization is 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 typically have negative autocorrelation, that is individuals typically exhibit a bias *against* repeating the same choice (3). This phenomenon is often referred to as the "Law of Small Numbers" (as subjects may try to reproduce in small sequences the properties of large sequences). The only possible exception is 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. As Camerer (1995) remarks, «whether they do in other settings, under natural conditions, is an empirical question.»

A simple way to test for randomness is to use the standard «runs test». Consider the sequence of strategies chosen by a player in the order in which they occurred $s = s_1, s_2, s_3 \ldots, s_n$, where , $S_X \in \{L, R\}$, and $n = n_L + n_r$ are the number of natural side and non-natural side choices made by the player. Let *f* (*r*; s) denote the total number of runs in the sequence *s*. A run is defined as a succession of one or more identical symbols that are followed and preceded by a different symbol or no symbol at all. Let *f*(*r; s*) denote the probability that there are exactly *r* runs in the sequence *s*. Let θ [*r; s*] = $\sum_{k=1,\ldots,r}$ *f*(*r; s*) denote the probability of obtaining *r* 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, the variable

$$\frac{r - 0.5 - 2\left(\dfrac{n_L n_R}{n}\right)}{\left[2n_L n_R\left(\dfrac{2n_L n_R - n}{n^2\left(n-1\right)}\right)\right]^2}$$

is distributed as a standardized Normal probability distribution *N*(0,1). The null hypothesis will then be rejected at the five-percent confidence level if the probability of *r* or fewer runs is less than .025 or if the probability of *r* or more runs is less than .025, that is if θ [*r; s*] < 0.025 or if 1 − θ [*r*-1; *s*] < 0.025. Similarly, at the ten-percent level, the hypothesis is rejected if they are less than 0.05 (4).

The results in Table 1 (in the next page) show the results of the Pearson test and the Runs test for 40 world-class soccer players, half kickers and half goalkeepers.

The null hypothesis of equality of payoffs cannot be rejected for the majority of players. It is rejected for just two players (David Villa and Frank Lampard) at the five-percent significance level and four players at the ten-percent significance level (in addition to Villa and Lampard, Iker Casillas and Morgan De Sanctis). Note that we should expect some rejections, just as if we flip 40 coins 10 times each we should expect some coins, but not many, to yield by pure chance proportions that are far from 50-50 such as 9 heads and 1 tail, or 8 heads and 2 tails. The confidence levels we are willing to adopt (typically no greater than five or ten percent) tell us how many rejections we should expect. In our case, with 40 players the expected number of rejections at the five-percent level is 0.05 x 40 = 2 and at the ten-percent level it is 0.10 x 40 = 4.

Thus, the evidence indicates that the hypothesis that scoring probabilities are identical across strategies cannot be rejected at the individual level for most players at conventional significance levels. The number of rejections is, in fact, identical to the theoretical predictions.

Behavior at the aggregate level also appears to be very consistent with equilibrium play. The joint hypothesis that each case is simultaneously generated by equilibrium play can be tested computing the aggregate Pearson statistic (summing up the individual Pearson statistics) and checking if it is distributed as a X^2 with *N* degrees of freedom. The results show that the Pearson statistic is 36.535 and its associated *p*-value is 0.627 for all 40 players. Hence, the hypothesis of equality of winning probabilities cannot be rejected at the aggregate level. Focusing only on kickers, the relevant statistic is 20.96 with a *p*-value of .399, and for goalkeepers it is 15.58 with a *p*-value of 0.742. Hence, the hypothesis of equality of winning probabilities cannot be rejected for either subgroup.

With respect to the null hypothesis of randomness, the runs tests show that this hypothesis cannot be rejected for the majority of players. They neither appear to switch strategies too often or too infrequently, but just about the right amount. This hypothesis is in fact rejected for just three players (David Villa, Alvaro Negredo and Edwin Van der Sar) and four players (in addition, Jens Lehman) at the five-percent and ten-percent significance levels. For the same reasons as in the previous test, we should be expecting two and four rejections.

Problems with the Runs Test. The runs test is simple and intuitive. However, it is a test that has low power to identify a lack of randomness. Put differently, current choices may be explained, at least in part, by past variables such as past choices or past outcomes, or past choices of the opponent, or interactions among these variables, and still the number of runs in the series of choices may appear to be neither too high or too low. In other words, many potential sources of dynamic dependence cannot be detected with a runs test. For this reason, some researchers on randomization have studied whether that past choices or outcomes have any role in determining current choices by estimating a logit equation for each player. For instance, in Brown and Rosenthal (1990) the dependent variable is a dichotomous indicator of the current choice of strategy, and 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 show that in fact it is possible to detect a number of dynamic dependences with this logit equation that are not possible to detect with the runs test (5).

Problems with Standard Logit Equation. Unfortunately, the standard logit equation is highly problematic in

TABLE 1
PEARSON AND RUNS TESTS

	Proportions			Success Rate		Pearson Test			Runs Test	
Name	#Obs	L	R	L	R	Statistic	p-value	r	–r[r-1,s]	–r[r,s]
Kickers:										
Alessandro Del Piero	55	0.345	0.654	0.736	0.805	0.344	0.557	24	0.237	0.339
Zinedine Zidane	61	0.377	0.622	0.782	0.815	0.099	0.752	26	0.126	0.192
Lionel Messi	45	0.377	0.622	1.000	0.928	1.270	0.259	22	0.416	0.544
Christiano Ronaldo	51	0.372	0.627	0.842	0.718	1.008	0.315	24	0.342	0.458
Mikel Arteta	53	0.433	a0.566	0.782	0.833	0.218	0.639	27	0.439	0.551
Xabi Prieto	37	0.324	0.675	0.833	0.880	0.151	0.697	16	0.256	0.392
Thierry Henry	44	0.477	0.522	0.809	0.782	0.048	0.825	19	0.086	0.145
Francesco Totti	47	0.489	0.510	0.782	0.833	0.195	0.658	20	0.070	0.119
Andrea Pirlo	39	0.384	0.615	0.733	0.833	0.566	0.451	20	0.505	0.639
Steven Gerrard	50	0.340	0.660	0.823	0.909	0.777	0.377	23	0.382	0.507
Samuel E'too	62	0.419	0.580	0.769	0.805	0.120	0.728	28	0.165	0.239
Diego Forlán	62	0.419	0.580	0.769	0.805	0.120	0.728	30	0.327	0.427
Roberto Soldado	40	0.400	0.600	0.937	0.750	2.337	0.126	21	0.539	0.667
Franc Ribéry	38	0.500	0.500	0.789	0.736	0.145	0.702	25	0.930	0.964
David Villa	52	0.365	0.634	0.631	0.909	5.978	0.014**	18	0.010	0.022**
Alvaro Negredo	45	0.288	0.711	0.769	0.906	1.501	0.220	26	0.986**	0.995
Ronaldinho	46	0.456	0.543	0.952	0.880	0.753	0.385	24	0.460	0.580
Martin Palermo	55	0.381	0.618	0.714	0.735	0.028	0.865	23	0.098	0.158
Frank Lampard	38	0.236	0.763	0.666	0.793	4.113	0.042**	17	0.791	0.898
Robbie Keane	42	0.309	0.690	0.769	0.758	1.174	0.278	17	0.184	0.296
All	962	0.386	0.613	0.795	0.822	20.96	0.3997			
Goalkeepers:										
Peter Cech	82	0.414	0.585	0.235	0.187	0.276	0.590	38	0.224	0.298
Víctor Valdes	71	0.394	0.605	0.214	0.232	0.032	0.857	32	0.196	0.272
Bodo Illgner	68	0.352	0.647	0.250	0.272	0.041	0.839	33	0.547	0.650
David James	69	0.391	0.608	0.185	0.238	0.270	0.603	40	0.924	0.954
Jens Lehman	72	0.444	0.555	0.218	0.225	0.004	0.949	28	0.014	0.026*
Edwin Van der Sar	80	0.412	0.587	0.121	0.148	0.125	0.722	26	0.000	0.001**
Mark Schwarzer	55	0.381	0.618	0.238	0.264	0.048	0.825	31	0.846	0.904
Oliver Kahn	58	0.379	0.620	0.227	0.138	0.747	0.387	33	0.881	0.928
Willie Caballero	60	0.350	0.650	0.095	0.230	1.674	0.195	29	0.522	0.634
Andreas Kopke	70	0.428	0.571	0.233	0.150	0.787	0.374	31	0.119	0.175
Tim Howard	67	0.402	0.597	0.222	0.225	0.000	0.978	30	0.169	0.241
Morgan De Sanctis	62	0.435	0.564	0.148	0.342	3.018	0.082*	34	0.700	0.783
Gorka Iraizoz	73	0.424	0.575	0.129	0.142	0.028	0.865	32	0.106	0.157
Gianluigi Buffon	71	0.408	0.591	0.241	0.142	1.113	0.291	35	0.420	0.518
Iker Casillas	69	0.347	0.652	0.250	0.088	3.278	0.070*	32	0.414	0.520
Julio Cesar	68	0.308	0.691	0.238	0.106	2.007	0.156	34	0.840	0.900
Andrés Palop	66	0.439	0.560	0.206	0.297	0.694	0.404	34	0.498	0.597
Pepe Reina	55	0.418	0.581	0.173	0.187	0.016	0.897	31	0.778	0.852
Stefano Sorrentino	48	0.458	0.541	0.136	0.269	1.275	0.258	27	0.687	0.783
Dani Aranzubia	68	0.455	0.544	0.225	0.189	0.138	0.709	29	0.062	0.098
All	1332	0.402	0.597	0.199	0.198	15.580	0.742			

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

SOURCE: Palacios-Huerta (2014).

that it generates *biased* estimates. Unawareness of this aspect is a common mistake in the literature (see Walker and Wooders (2001), Wooders (2010), Van Essen and Wooders (2015). The basic issue is that the choice of strategy in a penalty kick may depend on certain observed characteristics of the player and his opponent, the specific sequence of past choices and past outcomes, and perhaps other variables. It may also depend on unobserved characteristics. Thus, the basic econometric problem is to estimate a binary choice model with lagged endogenous variables and unobserved heterogeneity where the effect of state dependence needs to be controlled for appropriately. The econometric estimation of these models is subject to a number of technical difficulties as parameter estimates jointly estimated with individual fixed effects can be seriously *biased* and *inconsistent*. Arellano and Honoré (2001) offer an excellent review of the issues that are encountered and how they can be resolved.

To establish that past choices have no significant role in determining current choices, it is possible to estimate a logit equation for each player using the Arellano and Carrasco (2003) method. The model generates *unbiased* and *consistent* estimates, allows for unobserved heterogeneity and for individual effects to be correlated with the explanatory variables.

Using this method it can be shown that the null hypothesis of randomization that all the explanatory variables are jointly statistically insignificant cannot be rejected for any player at the five-percent level, and is rejected for only three players (David Villa, Frank Lampard, and Iker Casillas) at the ten-percent level.

The main finding in this section is that 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 possible strategic links between subsequent penalty kicks. These results, which extend Palacios-Huerta (2003), represent the first time that both implications of von Neumann's (1928) Minimax Theorem are supported in real life.

LAB AND FIELD EXPERIMENTAL TESTING

General Considerations. Although Vernon Smith received the 2002 Nobel Prize in Economic Sciences «for having established laboratory experiments as a tool in empirical economic analysis», this tool has become under severe attack in recent years. A main critique is that the data generated in laboratory experiments are not «realistic», and hence to obtain more realistic data we should pursue experiments not in the lab but in the field.

Falk and Heckman (2009) explain in some detail why this critique is not only misguided but plain wrong. Consider an outcome of interest Y and a list of determinants $X_1, \ldots, X_N$. Suppose that:

$$Y = f(X_1, X_2, \ldots, X_N)$$

Now we are interested in knowing the causal effect of X_1 on Y, that is the effect of varying X_1 holding fixed $X^* = (X_2, \ldots, X_N)$. Thus, unless f is *separable* in X_1, so that $Y = \theta(X_1) + g(X^*)$, the level of Y response to X_1 will depend on the level of X^*.

Further, even in this separable case, unless $\theta(X_1)$ is a *linear function* of X_1, the causal effect of X_1 will depend on the level of X_1 and the size of the variation of X_1. These are problems that appear *both* in field and lab experiments, and in any estimation of the causal effect of X_1.

X^* may be demographic characteristics, individual preference parameters, social influences, or any set of aspects of the environment. Let X^* denote all these characteristics in a lab setting (say with student subjects), and X^{**} denote these characteristics in a natural setting (say, with sportscards traders as subjects). If one is interested in the causal effect of X_1 on Y, which one is more informative: holding fixed X^* or holding fixed X^{**}?

Well, experiments are able to obtain universally defined causal effects of X_1 on Y:

– only under assumption $Y = \theta(X_1) + g(X^*)$,

and

– only if the response of Y to X_1 is linear.

But if this is the case, then lab experiments and field experiments are *equally* able to obtain accurate inferences about universal effects. Therefore, the general quest for running experiments in the field to obtain more «realistic» data is fundamentally misguided. In other words, if the exact question being asked and the population being studied are exactly mirrored in an experiment, then the information from the experiment can be clear and informative. In fact, it should be identical.

Camerer (2011) reviews the available studies on markets, student donations, fishing, grading, sports cards and restaurant spending that provide the closest matches of lab and field settings, protocols and subjects and confirms these predictions. He concludes that «no replicated evidence that experimental lab data fail to generalize to central empirical features of field data (when the lab features are deliberately closely matched to the field features). … The default assumption in the economics profession should be that lab experiments are likely to generalize to closely matched field settings. … This is the default assumption, and is generally supported by direct comparisons, in other fields such as biology studies comparing animal behavior in lab settings and in the wild».

With this idea in mind, it is possible to study in an experimental lab if the *same* professional football players play the *same* formal game according to the Minimax predictions as well. Palacios-Huerta (2014) recruited subjects come from clubs in the north of Spain, a region with a high density of professional teams professional soccer clubs. Eighty male professional soccer players (40 kickers and 40 goalkeepers) were recruited to form forty kicker-goalkeeper pairs. Subjects who previously played for the same team were not allowed to participate in the same pair. This measure was implemented to parallel the reality that players encounter in the field, as we would not want friends or former teammates to play this game against one another. The results he finds is that, as predicted, professional football players also play Minimax in the lab.

Thus we learn than when the exact question being asked is mirrored in a laboratory experiment, and the population being studied is the same as in the field, the outcomes from the experiment can be just as clear and informative. In fact, that can be essentially identical.

TABLE 2
PPROPORTION OF PROFESSIONAL PLAYERS THAT INDICATE THAT THEY WOULD ALWAYS KICK TO THE SAME SIDE IN A REAL LIFE PENALTY KICK

US*	England	France	Spain	Germany
44% (N = 20)	0% (N = 24)	0% (N = 300)	0% (N = 42)	0% (N = 57)

* Note: US data comes from the survey reported in the working paper version of Levitt, List and Reiley (2010).

This suggests that when either the exact question being asked is not mirrored or the population being studied differs, the outcomes from the experiment will probably not parallel those observed in the field.

As such it is easy *not* to obtain Minimax behavior in the lab. Consider for example players who *do not play* Minimax in the field and who say, when asked, that they *would not play Minimax* in real life: players who play in the Major League Soccer in the US. These are the players studied in Levitt, List and Reiley (2010). Table 2 below makes it clear that the US subjects are drastically different from their European counterparts who never indicate that they would never randomize, whereas among US players more than 40% say they would always kick to the same place.

Levitt, List and Reiley (2010) not only study US players but they also form pairs of friends that know each other (teammates) for their experimental tests. But we know that *friends* do not behave like *enemies*. The game friends play is certainly embedded in a larger game not captured by a zero-sum experimental design. Since friends cooperate, the very control of having friends play against one another in a lab experiment is an artificial margin that causes deviations from equilibrium in a game that requires strict competition.

Laboratory studies of strictly competitive games (and possibly all other games) should benefit from capturing the fundamental competitive conditions that subjects encounter in real life. Not capturing these conditions, by construction, definitely induces behavior different than that observed in real life. As a result, and not surprisingly, Levitt, List and Reiley (2010) find that their US players do not play Minimax in the locker room (which is what they used as a «lab setting}).

Thus the results testify that what does not happen in the field, should not and does not happen in the lab.

NEUROECONOMIC EVIDENCE ON STRICTLY COMPETITIVE STRATEGIC SITUATIONS

Over the last couple of decades, a new field called Neuroeconomics is being developed with the objective «to create a theory of economic choice and exchange that is neurally detailed, mathematically accurate, and behaviorally relevant» (Camerer, 2008). This section reports measures of neural activity in the two dimensions that characterize Minimax when subjects are playing a zero-sum game almost identical to the penalty kick game described earlier.

The study whose results were reported in Palacios-Huerta, Olivero, Bestmann, Vila and Apesteguia (2014) was performed in the Hospital Nacional de Parapléjicos de Toledo during 2012 with a total of 20 healthy subjects. They formed 20 pairs: 20 volunteers were studied inside an fMRI scanner and another twenty subjects were studied outside the scanner. The two subjects were not friends and had not met before. One player was playing in a computer in a quiet room located outside the scanner room. The other player was lying down within the MRI room. For this player, the PC monitor was substituted by MRI compatible goggles and the keyboard was substituted by a button box designed for the hand. Player location (inside or outside the MRI) was decided by flipping a coin. They played this game:

	A	*B*
A	60, 40	95, 5
B	90, 10	70, 30

It was found that the results of the Pearson tests of equality of payoffs across A and B strategies, as well as the results of the runs tests, conformed quite close to Minimax. They are even an order of magnitude closer to Minimax than the US players' behavior in the locker room observed in Levitt, List and Reiley (2010).

From a neuroeconomics perspective, the fMRI revealed activity increases in various bilateral prefrontal regions during the decision period. Interestingly, activity in left inferior prefrontal cortex related significantly to the ability to equate payoffs. In other words, in this prefrontal region correlated with the performance measure for equating payoffs, with higher activity in participants who more effectively succeeded in equating payoffs. Conversely, a contralateral, right inferior prefrontal region related to the ability to generate random sequences of choices. Activity in these regions was correlated with the performance score testing for the randomness of choices using the *p*-value of the runs test.

Together these data suggest that two inferior prefrontal nodes jointly contribute to the ability to optimally behavior in strictly competitive strategic situations. It is not known what the future will bring in terms of our capabilities to observe the workings of the brain and make inferences, but this evidence shows for the first time the parts of the brain where strict strategic competition takes place.

RECENT FINDINGS

Finally, we conclude with a brief description of some recent findings of interest.

It is known that subjects can detect and exploit non-equilibrium play in zero-sum games with unique equilibrium in mixed strategies (Shachat and Swarthout, 2004). In a recent experimental game theoretical study, Gil and Prowse (2016) find striking differences according to the cognitive ability of subjects: more cognitively able subjects behave closer to equilibrium, converge more frequently to equilibrium play and earn more even as behavior approaches the equilibrium prediction. The average level of more cognitively able subjects responds positively to the cognitive ability of their opponents, while the average level of less cognitively able subjects does not respond. A consequence of these findings is that to the extent that experiences contributes to cognition (and also forms appropriate competitive character skills), experience becomes a powerful predictor of equilibrium behavior in the lab.

Green, Rao and Rothschild (2016) observe experts perform a task in the lab that is logically isomorphic to a familiar task in which they are skilled. They find that performance plummets when contextual cues disappear, implying that the expertise we observe on the familiar task in the field will travel far in the lab if the relevant contextual cues are maintained. They conclude that «this observation entails a normative implication for experimental design. If economic actors approximate rationality though context-dependent heuristics, then lab studies which abstract away contextual cues bias their findings against standard theories of rationality». Failing to maintain the contextual cues explains the results in Levitt, List and Reiley (2010), plus of course the fact that their players do not play Minimax in the field. See Palacios-Huerta (2014, chapters 2-3).

Erev, Roth and Slonim (2016) reports on an experimental design to evaluate how well the Minimax hypothesis describes behavior across a population of games by randomly sampling constant sum games with two players and two actions with a unique equilibrium in mixed strategies. They find that students behavior is more consistent with Minimax play the closer the mixed strategy equilibrium is to equal probability play of each action.

Randomization, a critical ingredient of Minimax, is also important in models of choice. Agranov and Ortoleva (2016) report experimental results in which they show that a majority of subjects *do exhibit* stochastic choices when presented repeteadly with the same question. This is in line with the interpretations of stochastic choice as emerging from an explicit preference for randomizing in human behavior, as opposed to emerging from random utility or mistakes.

NOTES

[1] Financial support from the Spanish Ministerio de Economía y Competitividad and FEDER (project ECO2015-66027-P) and from the Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (IT-869-13) is gratefully acknowledged. This survey is partly based on previous published research that appeared in my book Beautiful Game Theory, Princeton University Press (2014), although it differs, sometimes markedly, from the original publication. Address: London School of Economics, Houghton Street, London WC2A 2AE, UK. Email: i.palacioshuerta@gmail.com Web: www.palacios-huer ta.com

[2] It could perhaps be modelled as a 3x3 game as well. Chiappori et al. (2000), however, study the aggregate predictions of this zero-sum game, rather than the Minimax individual predictions, and conclude that the availability of C as an action is not an issue. Their findings are also substantiated in the dataset discussed in this survey. This means that a penalty kick may be basically described as a two-action game.

[3] See Bar-Hillel and Wagenaar (1991), Rapoport and Budescu (1992), Rapoport and Boebel (1992), and Mookherjee and Sopher (1994). Neuringer (2002), Rabin (2002) and Camerer (1995) review the literature. See also Tversky and Kahneman (1971).

[41] Aggregate level tests may also be implemented by checking if the values in columns □ [r; s] and □ [r-1; s] tend to be uniformly distributed in the interval [0, 1], which is what should happen under the null hypothesis of randomization. See Palacios-Huerta (2003, 2014).

[5] Compare Table IV in Brown and Rosenthal (1990) with Table 4 in Walker and Wooders (2001). There are many subjects that pass the runs test but do exhibit serial dependence in that a number of lagged endogenous variables (choices and outcomes) help predict their subsequent choices.

REFERENCES

AGRANOV, M. and ORTOLEVA, P (2016). «Stochastic Choice and Preferences for Randomization». *Journal of Political Economy*, forthcoming.

ARELLANO, M. and HONORÉ, B. (2001). «Panel Data Models: Some Recent Developments». In *Handbook of Econometrics*, vol. 5, chapter 53, ed. James E. Heckman and Edward Leamer, 3229-3296. North-Holland.

ARELLANO, M. and CARRASCO, R. (2003). «Binary Choice Panel Data Models with Predetermined Variables». *Journal of Econometrics*, nº 115, pp. 125-157.

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, pp. 460-482.

BROWN, J. and ROSENTHAL, R. (1990). «Testing the Minimax Hypothesis: A Reexamination of O'Neill's Experiment», *Econometrica*, nº 58, pp. 1065-1081.

CAMERER, C.F. (2011). «The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List». Caltech, Mimeo.

CAMERER, C.F. (2003). *Behavioral Game Theory*, Princeton University Press.

CHIAPPORI, P.A.; LEVITT, S.D. and GROSECLOSE (2002), T. «Testing Mixed Strategy Equilibrium When Players Are Heterogeneous: The Case of Penalty Kicks». *American Economic Review*, nº 92, pp. 1138-1151.

EREV, I.; ALVIN E.R. and SLONIM, R. (2016). «Minimax across a population of games». *Journal of the Economic Science Association*, vol. 2, pp. 144-156

FALK, A. and HECKMAN, J. (2009). «Lab Experiments Are a Major Source of Knowledge in the Social Sciences». *Science* 326(5952): pp. 535-538.

GREEN, E.; JUSTIN, M.; RAO and DAVID M. ROTHSCHILD (2016). «A Sharp Test of the Portability of Expertise»., SSRN Working Paper, September.

GILL, D. and Prowse, V. (2016). «Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A level-*k* analysis», *Journal of Political Economy* 2016, vol. 124, nº.6, pp. 1619-1676.

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

LEVITT, S.D.; LIST, J.A. and REILEY, D. (2010). «What Happens in the Field Stays in the Field: Professionals Do Not Play Minimax in Laboratory Experiments». *Econometrica, vol.* 78, nº 4, pp. 1413-1434.

NASH, J.F. (1950). «Equilibrium Points in *N*-person Games».*Proceedings of the National Academy of Sciences* vol. 36, nº 36, pp. 48-9.

NASH, J.F. (1951). «Non-Cooperative Games». *Annals of Mathematics, vol.* 54 nº 54, pp. 286-95

NEURINGER, A.J. (1986). «Can People Behave 'Randomly'?: The Role of Feedback». *Journal of Experimental Psychology: Genera*l, nº 115, pp. 62-75.

O'NEILL, B. (1991). «Comments on Brown and Rosenthal's Reexamination». *Econometrica*, nº 59, pp. 503-507.

PALACIOS-HUERTA, I. (2003). «Professionals Play Minimax». *Review of Economic Studies*, nº 70, pp. 395-415.

PALACIOS-HUERTA, I. (2014). *Beautiful Game Theory. How Soccer Can Help Economics.* Princeton University Press, NJ.

PALACIOS-HUERTA, I. and VOLIJ, O. (2008). «Experientia Docet: Professionals Play Minimax in Laboratory Experiments». *Econometrica*, vol. 76, nº 1, January, pp. 71-115.

PALACIOS-HUERTA, I. OLIVERO, A.; BESTMANN, S.; VILA, J.F. and APESTEGUIA, J. (2014). «Mapping Minimax in the Brain». Chapter 4 in Palacios-Huerta, I. *Beautiful Game Theory*, pp. 58-67.

RABIN, M. (2002). «Inference By Believers in the Law of Small Numbers». *Quarterly Journal of Economics*, nº 117, pp. 775-816.

RAPOPORT, A. and BOEBEL, R.B. (1992). «Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis». *Games and Economic Behavior*, nº 4, pp. 261-283.

RAPOPORT, A. and Budescu, D.V. (1992). «Generation of Random Sequences in Two-Person Strictly Competitive Games«. *Journal of Experimental Psychology: General*, nº 121, pp. 352-363.

RUBINSTEIN, A. (1991), «Comments on the Interpretation of Game Theory». *Econometrica*, vol. 59, nº. 4, July 1991, pp. 909-924.

SCHELLING, T. (1960). *The Strategy of Conflict*. Harvard University Press.

SCHWEIZER, M. (2016). «Minimax Play Around the World: Do Tennis Players Play Optimally?». London School of Economics, Master Thesis, May.

SHACHAT, J. and J. SWARTHOUT, T. (2004).«Do we detect and exploit mixed strategy play by opponents?». *Mathematical Methods of Operations Research,* July, Volume 59, Issue 3, pp 359-373

SPILIOPOULOS, L. (2016). «Minimax Play at Wimbledon Redux». *SSRN Working Paper*.

TVERSKY, A. and KAHNEMAN, D. (1971). «The Belief in the Law of Small Numbers». *Psychological Bulletin*, nº 76, pp. 105-110.

VAN ESSEN, M.; WOODERS, J. (2015). «Blind Stealing: Experience and Expertise in a Mixed-Strategy Poker Experiment». *Games and Economic Behavior* nº 91, pp. 186-206

VON NEUMANN, J. (1928). «Zur Theorie der Gesellschaftsspiele». *Mathematische Annalen*, nº 100, pp. 295-300.

VON NEUMANN, J. (1953). «Communication on the Borel Notes». *Econometrica,* vol. 21, No. 1 (January), pp. 124-127.

WALKER, M. and WOODERS, J. (2001). «Minimax Play at Wimbledon». *American Economic Review*, nº 91, pp. 1521-1538.

WOODERS, J. (2010). «Does Experience Teach? Professionals and Minimax Play in the Lab». *Econometrica,* nº 78, pp. 1143-1154.

WANG, Z. (2009). «Mixed) Strategy in Oligopoly Pricing: Evidence from Gasoline Price Cycles Before and Under a Timing Regulation». *Journal of Political Economy*, vol. 117, nº.6, pp. 987-1