Do soccer players play the mixed-strategy Nash equilibrium?

Azar, Ofer H. and Bar-Eli, Michael

Ben-Gurion University of the Negev

2009

Do soccer players play the mixed-strategy Nash equilibrium?

Ofer H. Azar[1,*] and Michael Bar-Eli[1]

Forthcoming in *Applied Economics*

[1] Department of Business Administration, Guilford Glazer School of Business and Management, Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva 84105, Israel.

Running title: Do soccer players play the mixed-strategy Nash equilibrium?

Abstract

Mixed-strategy Nash equilibrium (MSNE) is a commonly-used solution concept in game-theoretic models in various fields in economics, management, and other disciplines, but the experimental results whether the MSNE predicts well actual play in games is mixed. Consequently, evidence for naturally-occurring games in which the MSNE predicts the outcome well is of great importance, as it can justify the vast use of MSNE in models. The game between the kicker and goalkeeper in soccer penalty kicks is a real-world game that can be used to examine the application of the MSNE concept or its accuracy because payoffs are a common knowledge, the players have huge incentives to play correctly, the game is simple enough to analyze, its Nash equilibrium is in mixed strategies, and players' actions can be observed. We collected and analyzed data on the direction of kicks and jumps in penalty kicks in various top leagues and tournaments. Our analysis suggests that the MSNE predictions are the closest to the actual sample data, even though some other prediction methods use information on the marginal distribution of kicks or jumps whereas the MSNE does not.

[*]Corresponding author. Tel.: +972-8-6472675; Fax: +972-8-6477691. E-mail address: azar@som.bgu.ac.il.

1. Introduction

Mixed-strategy Nash equilibrium (MSNE) is a common solution concept employed in many theoretical and applied-theory articles in economics, management, and other disciplines. In a pure-strategy Nash equilibrium, each player chooses an action and the actions constitute an equilibrium if given the equilibrium actions of the other players, no player finds it beneficial to deviate from his equilibrium action and choose another action instead. In a mixed-strategy Nash equilibrium, on the other hand, players adopt mixed strategies, which means that they randomize (not necessarily with equal probabilities) between several actions (they may randomize between all their possible actions, or only between a subset of them). In an MSNE, each player's mixed strategy is optimal given the equilibrium mixed strategies of the other players. In other words, no player has an incentive to deviate from his mixed strategy to another pure or mixed strategy, conditional on the other players choosing their equilibrium strategies.

One reason for the popularity and importance of the MSNE concept is that every finite strategic-form game has a mixed-strategy equilibrium, while it does not necessarily have a pure-strategy equilibrium (Nash, 1950). However, it is not clear why players should play the strategies dictated by the MSNE. It is hence important to examine whether in actual games players play according to the MSNE or not. This question has been addressed by several experimental studies, with mixed results.[1] Results that support MSNE play in experiments were obtained by O'Neill (1987), in an experiment involving a repeated two-person constant-sum game, but later it was argued that O'Neill's data support the conclusion of MSNE play less than argued by O'Neill (Brown and Rosenthal, 1990). Additional support of MSNE play was obtained by McCabe, Mukherji and Runkle (2000) in a three-person matching-pennies game played with perfect monitoring and complete payoff information.

---

[1] In order to be able to use a mixed strategy effectively, the player should be able to randomize his actions, and therefore the psychology literature that examines whether people can produce random series is also related to the question whether people play the MSNE. The interested reader is referred to Rapoport and Budescu (1992) and Budescu and Rapoport (1994) and the references they cite.

Somewhat more ambiguous results (regarding whether players play the MSNE or not) were obtained by Rapoport and Boebel (1992), who conducted two studies designed to investigate interactive behavior in two-person zero-sum games and assess the descriptive power of the minimax hypothesis. Additional ambiguous results are reported by Rapoport and Amaldoss (2000, 2004): in both articles, they find mixed-strategies equilibrium play on the aggregate but not the individual level.

Mookherjee and Sopher (1994) found support for equilibrium play in a matching-pennies game when subjects had only two possible actions and were provided with complete information regarding opponent's choices and payoffs. In a later article, however (Mookherjee and Sopher, 1997), they discovered that in constant-sum games with at least four choices available to each player, observed behavior departed significantly from the equilibrium predictions. Ochs (1995) also obtained results that deviated from the Nash equilibrium predictions, in an experiment studying the choices of subjects playing mixed extensions of three variants of simple 2X2 non-constant sum matching-pennies games. Erev and Roth (1998) examined learning in games with a unique equilibrium in mixed strategies, analyzing both the ex-ante and ex-post descriptive power of learning models, and found that even a one-parameter reinforcement learning model robustly outperforms the equilibrium predictions. Shachat (2002) introduces a new methodology for eliciting mixed strategies and finds evidence that subjects do not play according to their minimax strategies.

The mixed results obtained in lab experiments, the importance of learning (see for example Ochs, 1995; Roth and Erev, 1995; Erev and Roth, 1998), and the inherent problem of external validity of lab experiments (i.e. to what extent can we learn from play in artificial games in the lab about behavior in naturally-occurring games in the real world) have led in recent years to several attempts to examine whether players play the MSNE in natural contexts. By looking at how experts play games, we can bypass the complexities of learning (because those experts have already learned and practiced the games they play). They generally have significant incentives to play correctly, and because we examine directly play in natural contexts, there is no question of external validity.

It is not hard to think about contexts in which players may randomize among several actions in real contexts; for example, as Osborne and Rubinstein (1994, p. 37-38) point out, tax authorities may randomize regarding which taxpayer to audit, and taxpayers may randomize whether to report their income truthfully. The fact that we observe people randomizing, however, does not guarantee that the probabilities with which they randomize are close to the probabilities predicted by the MSNE. To examine whether people play the MSNE in natural contexts, we should find a natural game in which payoffs are common knowledge, the players have significant incentives to play correctly, the game is simple enough to analyze, its Nash equilibrium is in mixed strategies, and players' actions can be observed. It is hard to find such games because natural games tend to be complex, but one interesting game satisfying the above criteria quite well is the game between the goalkeeper and the kicker during penalty kicks in soccer.[2]

Penalty kicks in soccer occur either after certain offenses, or at the end of the game, to untie a game or a match, in certain tournaments (including the World Cup, the European Cup, and many other top tournaments). A kick is shot from the penalty mark which is located 11 meters from the goal, the goalkeeper is not allowed to move forward before the ball is kicked, and no player other than the goalkeeper is allowed to stand between the goal and the penalty mark, thus giving the kicker a very high probability of scoring a goal. Because of the short distance between the ball and

---

[2] Another sports context that allows to analyze whether professional players play the MSNE is tennis. Indeed, the first to examine MSNE play in sports are Walker and Wooders (2001), who looked at win rates in the serve and return play of top professional tennis players at Wimbledon. In addition, soccer provides opportunities to examine MSNE play not only in penalty kicks. Moschini (2004), for example, examined whether soccer players, during the regular game (not in penalty kicks), choose to kick to the near goalpost or the far one, concluding that their behavior is consistent with the MSNE. Finally, soccer has also served as a case study for additional economic studies, such as studies about the relationship between pay and performance (Torgler and Schmidt, 2007), the effect of soccer results on the stock market (Scholtens and Peenstra, forthcoming; Klein, Zwergel and Fock, forthcoming), the effect of hosting the soccer World Cup on unemployment (Hagn and Maennig, forthcoming), and the efficiency of betting markets (Forrest and Simmons, 2008).

the goal, and because of the high speed of the ball during penalty kicks[3], the goalkeeper generally cannot afford to wait until he sees clearly to which direction the ball is kicked; rather, he has to make a decision whether to jump to one of the sides or to stay in the center at about the same time that the kicker chooses where to direct the kick. This creates a simple but interesting real-life example of a game in which the Nash equilibrium is in mixed strategies. The reason that no pure-strategy equilibrium exists for this game is that if one of the players adopts a pure strategy, the other player can take advantage of this, and then the first player no longer finds it optimal to choose that pure strategy. For example, if the goalkeeper chooses the pure strategy of jumping left, meaning that he always jumps left, then the kicker will always kick to the other direction. But then the goalkeeper should deviate and not jump left, and therefore no equilibrium exists with pure strategies. When the players randomize between several actions we can find an equilibrium for this game (as is analyzed in more detail in section 3), and therefore an MSNE does exist in this game. In practice, the players do not completely randomize. The goalkeeper uses his knowledge of the directional distribution of penalty kicks in general, the past behavior of the kicker, and cues he might obtain from the kicker's behavior in the seconds before the kick, to try to decide correctly to which side to jump, if at all. Similarly, the kicker chooses where to direct the ball according to his conjecture about where the goalkeeper is likely to jump, where stopping the ball is harder, etc. Even though the players' choices are not completely random, we can treat these choices as coming from a mixed strategy of each player.

Because a goal achieved in a penalty kick counts as any other goal, and because the number of goals scored in an average soccer game is very small[4], the importance of the goalkeeper's (as well as the kicker's) performance during the penalty kick is tremendous. World Cup and European Cup games have been won many times by the penalty kicks that followed a tied game, for example. Not only the outcome of the game depends on the goalkeeper's performance during a penalty kick, but

---

[3] The time it takes the ball to reach the goal from the penalty mark is about 0.2-0.3 seconds (see Chiappori, Levitt and Groseclose, 2002; Palacios-Huerta, 2003).

[4] For example, in the Korea/Japan 2002 World Cup, an average of 2.52 goals per match were scored (see http://www.fifa.com/images/pdf/IP-301_12A_comparative.pdf).

also huge amounts of money are involved. Players receive bonuses for winning games, teams can make large amounts of money by winning and climbing to the next stage of a tournament, and the goalkeeper's reputation and thus future earnings also depend on his performance, to give a few examples. Since players in top leagues earn hundreds of thousands and often millions of dollars annually, it is clear that the goalkeeper's performance and reputation have a very significant monetary effect, and the goalkeeper has huge incentives to do his best in general and in the few seconds of a penalty kick in particular.

We examine whether the decisions made by kickers and goalkeepers during penalty kicks are close to the predictions of the MSNE. We collected data on penalty kicks in top leagues and championships worldwide (the data are described in more detail in the next section). While the players in the various penalty kicks are not always the same players, the reasons that lead players to use mixed strategies in many games (such as matching pennies) are still relevant, and therefore it is still appropriate to examine whether the players play the MSNE. For example, if a certain goalkeeper adopts a strategy of always jumping left, this strategy will soon be noticed and learned by kickers of other teams, who will direct penalty kicks to the other side when playing against this goalkeeper. Similarly, if one of the kickers in a team always kicks to the right, this strategy will soon be observed by others, and goalkeepers who face this kicker during penalty kicks will jump to the direction of the kick. Indeed, teams keep records of past behavior of players in other teams during penalty kicks. Consequently, the incentive to randomize exists for both the kickers and the goalkeepers.

When we started the project and collected the data, no other study of mixed-strategy play in penalty kicks was published, but recently two important contributions were added to the literature: Chiappori, Levitt and Groseclose (2002) (CLS), and Palacios-Huerta (2003) (PH). CLS built a theoretical model that describes the penalty-kick game as a simultaneous 3X3 game between the kicker and the goalkeeper, where each can choose left, center, and right.[5] They make certain assumptions about the payoffs in the game (e.g. that if both players choose "center," the goalkeeper always stops the ball), and these assumptions yield predictions about which strategies or combinations of strategies should be more common than others if players play the MSNE (e.g., the kicker chooses left more often than right). They find that these predictions hold in data they collected on penalty kicks in the French and Italian first leagues.

PH examines whether penalty kicks satisfy the assumptions of minimax play using a dataset of penalty kicks from various countries (mostly Italy, Spain and England). He performs most of the analysis on a simplified 2X2 game (the goalkeeper and the kicker can each choose only right or left) rather than a 3X3 game. He finds that the winning probabilities of each strategy of each player are similar, and that players' choices are serially independent.

While we also explore penalty kicks in soccer, there are significant differences between our article and those of CLS and PH. First, our analysis allows the goalkeeper and the kicker to choose also center in addition to left and right (as opposed to PH). The addition of the center makes a big difference, for two reasons. First, the payoffs to choosing center are substantially different from those to choosing right and left, whereas the differences between right and left are not that large. Second, in our analysis there are nine possible strategy combinations in each kick (3X3), whereas in the analysis of PH there are only four. Both of the above differences imply that from the players' perspective a 3X3 game is much more complex and it makes it more difficult for them to figure out how to play the MSNE. Finding that the MSNE still predicts reasonably well the outcomes in this more complex game therefore provides much more convincing evidence for the claim that players play the MSNE, and therefore is an important contribution over the existing literature.

Second, as opposed to CLS, we look at the predictions of the MSNE about the exact probabilities of the joint distribution of kicks and jumps. Finding that the actual probabilities of

---

[5] The reader who compares our results to theirs should notice that they define "right" as a kick to the right from the kicker's perspective, while we define "right" as the right from the goalkeeper's perspective (i.e. our directions are the opposite of theirs; PH uses the goalkeeper's perspective as we do).

different actions in the empirical data match closely the probabilities predicted by the MSNE provides strong support for the MSNE concept and is therefore an important addition to the current literature.

Finally, we explore various alternatives to the MSNE predictions, and show that the MSNE yields the best predictions. This approach significantly adds to the results in the former two studies, where the MSNE was not confronted with other alternatives to see whether it predicts better or not. One of the alternatives we consider is that the players play according to probability matching - playing each action with a probability that is identical to the probability that this action is optimal.[6] The rationale for examining this idea is that many experimental studies (for a literature review see Vulkan, 2000) show that probability matching is a very common behavior in situations that resemble the one of penalty kicks. To examine the predictions implied by probability matching without consulting the actual marginal distribution of kicks, we surveyed 21 of the best goalkeepers in Israel, and obtained their perception about the distribution of penalty kicks, in two different methods.

Given the vast use of the MSNE concept in theoretical models and the ambiguous results in experimental studies that test whether players play the MSNE, it becomes especially interesting and important to examine play of experienced players who have high incentives to perform optimally, in a natural context. We do so in the context of soccer penalty kicks and find that the MSNE predicts the behavior of soccer players better than various alternatives. This is an important support for the vast use of the MSNE concept to analyze games and to predict their outcomes.

The rest of the article is organized as follows: Section 2 describes the data collection process and presents the data. Section 3 analyzes the data to determine whether the observed behavior of kickers and goalkeepers is reasonably close to the mixed-strategy Nash equilibrium. The last section concludes.

[6] Notice that this is not optimal play; we elaborate on this later on.

2. The Data

In order to collect data on penalty kicks during real games, we searched in the archives of various television channels, found different soccer games in various leagues and championships worldwide, and watched the games to see whether they involved penalty kicks. For those penalty kicks that we found, we asked three independent judges to determine to which part of the goal the ball was kicked, to which direction the goalkeeper jumped (if at all), and whether he stopped the ball, using a diagram of the goal's area.[7] This process yielded a dataset of 286 penalty kicks.[8]

We can think about each penalty kick as a game in which the goalkeeper chooses one of three actions (staying at the center, jumping to the right, or jumping to the left) and the kicker chooses one of three actions (kicking to the center, right, or left). There is a potential confusion about directions in this context. If the goalkeeper jumps to his left, and the kicker kicks towards his left, these are two opposite directions even though both are "left." To avoid this confusion, in what follows, every time we mention right or left, it is from the goalkeeper's perspective. Table 1 presents the outcomes in the 286 observations, divided according to the direction of jumps and kicks.

---

[7] To get a feeling for the task the goalkeeper performs, we note that the size of a soccer goal is 8 yards (7.32 meters) wide and 8 feet (2.44 meters) high.

[8] Some additional details about the data collection process are reported in Bar-Eli et al. (2007), who combine these data with a survey of goalkeepers (not the one reported later in this article) in order to examine whether goalkeepers exhibit an action bias – having a preference to jump rather than to stay in the center. However, Bar-Eli et al. do not consider the implications of the data with respect to the Nash equilibrium as we do here. Instead, they take the kickers' actions as given and only ask whether goalkeepers seem to behave optimally. Interestingly, while they claim that goalkeepers seem to be biased in favor of jumping, we find here that the MSNE predicts the outcome very well. Notice, however, that Bar-Eli et al. examine the optimality of goalkeepers' behavior given the actual kickers' behavior, whereas the predicted behavior of goalkeepers in the MSNE implicitly assumes that the kickers play the MSNE as well (and kickers' actual play is not identical to kickers' MSNE play). A possible interpretation of the results is that the action bias is small in relation to the strength of the MSNE concept, and thus, despite the possible bias in goalkeepers' behavior, the MSNE still yields better predictions than other alternatives.

Table 1: Joint Distribution of Jumps and Kicks

		Goalkeeper			
		Left	Center	Right	Total
Kicker	Left	54 (18.9%)	1 (0.3%)	37 (12.9%)	92 (32.2%)
	Center	41 (14.3%)	10 (3.5%)	31 (10.8%)	82 (28.7%)
	Right	46 (16.1%)	7 (2.4%)	59 (20.6%)	112 (39.2%)
	Total	141 (49.3%)	18 (6.3%)	127 (44.4%)	286 (100%)

The next step is to determine the utility of each player from each outcome. In any specific kick either a goal is scored or not. When a goal is scored, the most natural payoffs to assign to the players are 1 to the kicker and -1 to the goalkeeper (i.e., the payoff represents the impact of the penalty kick on the score in the game). When a goal is not scored, the payoffs are 0 to both the goalkeeper and the kicker. This is illustrated in Table 2.

Table 2: Payoff Matrix

	Goal is scored	Goal is not scored
Kicker's payoff	1	0
Goalkeeper's payoff	-1	0

The game can have 9 different realizations of strategy choice by both players (3X3), and the payoffs that the players should consider when evaluating the game (and when we compute the MSNE) are the expected payoffs from each pair of actions by the two players. These can be obtained by looking at the percentage of goals scored for each pair of actions chosen by the players. Table 3 presents information about the kicks that were stopped by the goalkeepers.

Table 3: Stopped Kicks

		Goalkeeper			
		Left	Center	Right	Total
Kicker	Left	16	0	0	16
	Center	4	6	1	11
	Right	0	0	15	15
	Total	20	6	16	42

Not surprisingly, most of the kicks stopped occur when the goalkeeper chooses the same direction in which the ball was kicked. We can see, however, that in a few cases a goalkeeper who jumped to one of the sides was still able to stop a ball directed towards the center. It is clear why this is possible if we remember that a kick that is classified as "Center" need not be at the exact center, but rather at any point in cells 4, 5, or 6, so jumping to the left allows the goalkeeper to potentially stop a ball directed a little to the left of the center.

Based on the data in Tables 1 and 3 we can compute the average percentage of kicks in which a goal was scored for each of the 9 cells in the tables, giving us the expected payoffs of the players from each outcome of the game. Table 4 presents these expected payoffs, computed as the number of goals scored divided by the number of kicks (for each combination of kick direction and jump direction).

Table 4: Expected Payoffs

		Goalkeeper		
		Left	Center	Right
Kicker	Left	0.704, -0.704	1, -1	1, -1
	Center	0.902, -0.902	0.4, -0.4	0.968, -0.968
	Right	1, -1	1, -1	0.746, -0.746

3. Do Goalkeepers and Kickers Play the Mixed-Strategy Nash Equilibrium?

Based on the payoffs presented in Table 4, we can examine whether the observed distribution of kicks and jumps is close to the distribution that results from the MSNE. In a mixed-strategy equilibrium in which each player plays every possible action with positive probability, each player must be indifferent between his possible actions. Let us denote the probability that the goalkeeper jumps to the left by p, and the probability that he stays in the center by q. Let us also denote the probability that the kicker kicks towards the left by α and the probability he kicks to the center by β. The expected utility from the three actions possible for each player, given the mixed-strategy adopted by the other player, is denoted as U (left), U (center), and U (right).

The kicker is indifferent between choosing left, center or right if and only if the following two equations hold:

$$U\ (\text{left}) = 0.704p + q + (1 - p - q) = 0.902p + 0.4q + 0.968(1 - p - q) = U\ (\text{center}), \text{ and}$$

$$U\ (\text{center}) = 0.902p + 0.4q + 0.968(1 - p - q) = p + q + 0.746(1 - p - q) = U\ (\text{right}).$$

Similarly, the goalkeeper is indifferent between choosing left, center or right if and only if the following two equations hold:

$$U\ (\text{left}) = -0.704\alpha - 0.902\beta - (1 - \alpha - \beta) = -\alpha - 0.4\beta - (1 - \alpha - \beta) = U\ (\text{center}), \text{ and}$$

$$U\ (\text{center}) = -\alpha - 0.4\beta - (1 - \alpha - \beta) = -\alpha - 0.968\beta - 0.746(1 - \alpha - \beta) = U\ (\text{right}).$$

Solving these two systems of equations yields the following values:[9]

$$p = 0.411,\ q = 0.110,\ \alpha = 0.344, \text{ and } \beta = 0.203.$$

If each player plays his mixed strategy according to these probabilities, independently of what the other does, it is simple to compute the expected frequency in each cell of the 3X3 table of actions. For example, the probability that we will observe an outcome in which both players choose "left" is given by $\alpha p = 0.344*0.411 = 0.141$. If we multiply the resulting nine probabilities by the number of kicks in the sample (286), we obtain a prediction about the distribution of kicks and jumps if kickers and players play simultaneously and play according to the MSNE. Table 5 presents this predicted distribution:

[9] We present here the values rounded to the third number after the decimal point, but carry on the computations below with the accurate values.

Table 5: Predicted Distribution According to the MSNE

		Goalkeeper			
		Left	Center	Right	Total
Kicker	Left	40.42	10.83	47.10	98.34
	Center	23.83	6.38	27.77	57.99
	Right	53.29	14.27	62.10	129.67
	Total	117.54	31.48	136.98	286.00

Not surprisingly, this table is not exactly equal to the actual distribution (see Table 1). The real question, however, is not whether the MSNE predictions are identical to the actual distribution. In a large sample of 286 different penalty kicks, and in a natural context (as opposed to a lab environment), where so many factors play a role in the decisions of the players, it will be naïve to expect that any method can be very accurate in predicting the actual distribution of kicks and jumps.

A more meaningful question to ask is whether the MSNE predictions are closer to the actual distribution than predictions obtained from other reasonable conjectures about how the players may play this game. The MSNE concept is a useful one if it predicts behavior better than other prediction methods that use the same information, even if its predictions are not completely accurate.[10] To compare the MSNE predictions to competing alternatives, we should do two things: one is to define a measurement method that allows us to quantify the difference between the predictions and the actual distribution, and thus to compare between the predictions obtained from different conjectures about how the players play the game. The second thing we should do is to formulate other reasonable conjectures about the players' strategies.

[10] An analogous example is the use of expected utility theory. Suppose that we obtained the exact utility function over all possible goods of hundreds of individuals. We can then compute the bundle they should buy according to expected utility theory, and compare it to the bundle they actually decide to purchase. If we have enough data, probably we can reject the hypothesis that people behave exactly according to expected utility theory. But this does not mean we have to dismiss this theory; the important question is whether expected utility theory predicts behavior better than competing theories. If it does, it is a useful theory, even if its predictions are not completely accurate.

To measure the difference between the predicted and actual distributions we employ two natural criteria. To proceed, let us first code the strategies left, center, and right as strategies 1, 2, and 3. Denote the actual number of penalty kicks where the kicker chose strategy i and the goalkeeper chose strategy j as A_{ij}, where i,j = 1, 2, or 3 (so A_{31} = 46, for example). Similarly, denote the respective predicted numbers as P_{ij} (for the MSNE, for example, P_{31} = 53.29). The first measure of inaccuracy between the actual and predicted distributions is the sum of the absolute values of the differences between the predicted and actual raw frequencies in each of the nine cells, which is denoted by AD (for "absolute differences"). Formally, for each prediction method, the AD value is equal to:

$$AD = \Sigma_i \Sigma_j |A_{ij} - P_{ij}|$$

The AD value can range from 0 (predictions are exactly equal to actual distribution) to 570.[11] The second criterion is similar, but takes the squared differences between the predicted and actual distributions instead of the absolute value, thus punishing for mistakes in prediction in a convex fashion (the marginal punishment being higher the higher is the mistake). This measure, denoted by SD (for "squared differences"), can range from 0 to 93,898, and is formally equal to:

$$SD = \Sigma_i \Sigma_j (A_{ij} - P_{ij})^2$$

Employing these two criteria on the MSNE predictions yields an AD value of 75.2, and an SD value of 817.0.

The next step is to formulate some other conjectures about how the players may play this game. One simple alternative is to examine the conjecture that the goalkeeper plays only after observing the direction of the ball (i.e. to replace the assumption that the game is simultaneous with an assumption that the kicker plays first). In this case, the payoff table (see Table 4) suggests that the goalkeeper should always jump to the direction of the kick (because then his chances of stopping it are the greatest, and therefore his expected utility is maximized). It is easy to see from Table 1 that

[11] The most inaccurate prediction is to predict that the kicker always kicks to the left and the goalkeeper always stays in the center; this prediction yields a predicted value of 286 in the cell where actually there is only 1 kick, and predicts 0 for all the other cells, which have together 285 actual kicks, resulting in the sum of absolute differences being 285X2 = 570.

this is not the case, but nevertheless the AD and SD values give us benchmarks to which we can compare the results obtained from other predictions. To employ this conjecture, we assume that the distribution of kicks is the actual one, and the goalkeeper jumps to the direction of the kick. This means that we predict that in 92 observations the kicker and goalkeeper both choose left, in 82 observations both choose center, and in 112 observations both choose right. The cells off the diagonal all have a predicted number of observations equal to zero. Notice that this method has an advantage over the MSNE: when using the MSNE, we did not use any data about the marginal distributions of kicks or jumps to predict their joint distribution, while here we use the actual marginal distribution of kicks to predict the joint distribution. Nevertheless, because the data are far from matching the conjecture that the goalkeeper can choose his action after observing the direction of the kick, the AD and SD values are high. The AD value is 326, and the SD value is 15,614.

Another simple conjecture is that because the ball is very fast during penalty kicks and the distance between the ball and the goal is only 11 meters, the goalkeeper has to choose to which direction to jump even before the ball is kicked, and the kicker can observe this before choosing the direction of the kick (i.e. the goalkeeper plays first). In this case, if the goalkeeper chooses right, the best for the kicker is to choose left, and vice versa. If the goalkeeper chooses center, the kicker is indifferent between right and left, so we will assume that he chooses each with equal probability.[12] In a similar fashion to what we did when we conjectured that kickers play first, we now use the actual marginal distribution of the jumps to predict the joint distribution of kicks and jumps. Table 6 presents the predicted joint distribution.

[12] Because the goalkeeper chooses center only in a few cases, other assumptions do not change much.

Table 6: Predicted Distribution if Goalkeeper Plays First

		Goalkeeper			
		Left	Center	Right	Total
Kicker	Left	0	9	127	136
	Center	0	0	0	0
	Right	141	9	0	150
	Total	141	18	127	286

Computing the AD gives a value of 390, and the SD is equal to 26,332. We can see that these values are higher than those of the conjecture that the kicker plays first, suggesting that to the extent that the game between the goalkeeper and the kicker is not completely simultaneous, goalkeepers might have better signals about kick direction than kickers have signals about jump direction. We can also see that so far the MSNE predictions are by far the closest to the actual data. This might be, however, not because the MSNE is a good prediction, but because the game is simultaneous and therefore the results obtained from assuming that one of the players plays first are far from the data.

Can we come up with a reasonable conjecture about how the players might play simultaneously, which will do better than the MSNE predictions? A natural alternative to the MSNE is to conjecture that the players randomize with equal probabilities among their three actions. Assuming that they play simultaneously, the randomization of the two players should be independent of each other, and consequently each of the nine cells in the table should have 286/9 = 31.78 observations. The AD value for this prediction is 156.2, and the SD value is 3585.6.

Another interesting conjecture worth exploring is that the goalkeepers or the kickers play according to probability matching, a well-known behavior that has been observed in many experiments (for a review of this literature, see Vulkan, 2000). To give a brief explanation of what probability matching means, assume that subjects have option A, which is optimal with a probability $p > 0.5$, and option B, which is optimal with a probability $(1-p)$. Also assume, for simplicity, that the difference in utility between the two options is constant; for example, one option wins a dollar and another wins nothing, but which option wins can change in each round of a repeated game or experiment, and the subjects do not know when making their choice in each round which option wins in that round. If subjects receive feedback after each round, they gradually learn the pattern of the game (i.e., the value of p), and then their optimal behavior is to always choose option A (because they have to choose before knowing which option is better in that round). It turns out, however, that people often choose A with probability p and B with probability (1-p). More generally, probability matching implies that when people have various possible actions, they choose each action with a probability that is equal to the probability that this action is the best response.

We can employ the probability matching principle either on the choices of the kickers or those of the goalkeepers. For the kickers, the optimal action is to kick left when the goalkeeper jumps to the right and vice versa, and to either kick left or right when the goalkeeper stays in the center (therefore we will assume that in the latter case the kicker chooses left and right with equal probabilities). According to the probability matching principle, the kickers should therefore kick left 47.6% of the kicks (the goalkeeper jumps to the right in 44.4% of the kicks, plus half of the 6.3% in which the goalkeeper stays in the center), and kick right 52.4% of the kicks. The goalkeepers are assumed to play according to the actual marginal distribution of jumps in the data. Because the game is simultaneous, the decisions are independent, and we can get the predicted distribution by multiplying the probabilities of the mixed strategies of the two players by each other, and then by the number of kicks (286). Notice that while this gives the same marginal distribution as when the goalkeepers play first (see Table 6), here the two players are playing simultaneously and therefore the joint distribution is not the same as in the former case. Kickers do not act optimally in each kick, as they did when goalkeepers played first. Table 7 presents the predicted joint distribution obtained:

Table 7: Predicted Distribution if Kickers Use Probability Matching

Kicker \ Goalkeeper	Left	Center	Right	Total
Left	67.05	8.56	60.39	136
Center	0	0	0	0
Right	73.95	9.44	66.61	150
Total	141	18	127	286

The AD value in this case is 164.0 and the SD value is 4361.7.

Alternatively, the goalkeepers might be the ones who use probability matching. In this case, they will choose to jump left with the same probability with which the kickers choose left, and similarly for center and right. This means that the marginal predicted distribution of both kicks and jumps will be equal to the actual marginal distribution of kicks. The predicted values in each cell, presented in Table 8, are obtained by multiplying the probabilities of the mixed-strategies of the two players by each other, and then by 286.

Table 8: Predicted Distribution if Goalkeepers Use Probability Matching (Using Actual Marginal Distribution of Kicks)

Kicker \ Goalkeeper	Left	Center	Right	Total
Left	29.59	26.38	36.03	92
Center	26.38	23.51	32.11	82
Right	36.03	32.11	43.86	112
Total	92	82	112	286

The AD value here is 130.2 and the SD value is 2597.4.

An interesting mix of the last two predictions is that each of the two players might play according to probability matching based on the actual marginal distributions. That is, goalkeepers choose directions with the same probabilities of the sample marginal distribution of kicks, and kickers choose the opposite direction of the marginal distribution of jumps (choosing a side randomly for the percentage of observations in which the goalkeeper stays at the center). This method results in the predicted distribution presented in Table 9:

Table 9: Predicted Distribution if Both Goalkeepers and Kickers Use Probability Matching

		Goalkeeper			
		Left	Center	Right	Total
Kicker	Left	43.75	38.99	53.26	136
	Center	0	0	0	0
	Right	48.25	43.01	58.74	150
	Total	92	82	112	286

The AD value obtained is 185.0 and the SD value is 5856.6.

We can see that so far the conjecture that goalkeepers use probability matching performs the best except for the MSNE. To try to improve it further, we wanted to see what happens if instead of using the actual marginal distribution of kicks we use the perception of goalkeepers about the distribution of kicks as the basis for the probability matching. This has two advantages: first, to the extent that goalkeepers perceptions are different from the actual distribution of kicks, if they use probability matching, it might be based on their perceived distribution rather than on the actual distribution.

Second, using the actual marginal distribution of kicks for the purpose of predicting the joint distribution of kicks and jumps in the probability matching case gives it an advantage over the MSNE, where we do not use information on the marginal distribution of kicks for prediction. By using the perceived distribution of goalkeepers rather than the actual distribution we alleviate this advantage. While it is likely that goalkeepers' perception of kick distribution is close to the actual distribution in general, which in turn is close to the distribution in our sample, there is still some difference between the perceived distribution and the sample distribution, and thus we do not use the exact marginal distribution of kicks in the sample to predict the joint distribution of kicks and jumps in the sample.

To elicit the perception of goalkeepers, we compiled a list of 69 goalkeepers who played in the three top leagues in Israel during the 2001-2 season. Then we chose four Israeli soccer experts, who were elite goalkeepers in the past and have remained involved in professional soccer to this day (as coaches etc.). We asked these four experts to rank each of the 69 goalkeepers on a 1-5 scale. We then asked the goalkeepers ranked 1-24 (according to the average ranking by the experts) to participate in a survey. 21 goalkeepers agreed, and we showed them a photograph of the goal area as it looks from the penalty kick mark. The goalkeepers were asked "Imagine to yourself that you are facing the best penalty-kicks kicker, and he kicks 10 penalty kicks. You are asked to mark 10 marks on the goal that represent the location to which the kicked ball arrives." We later converted their drawings to left, center, and right by dividing the goal area to three equal areas. As a robustness check, we also showed the goalkeepers a diagram of the goal divided to different areas and asked them "It is known that in one season there are on average 100 penalty kicks in the National League. You are asked to write in numbers how the kicks are distributed. Please check yourself that the numbers are accurate and their sum is 100." Table 10 presents the results, aggregated over the 21 goalkeepers:

Table 10: Perceived Distribution of Penalty Kicks Given by 21 Top Goalkeepers

	Left	Center	Right	Total
10 kicks of best player	88 (41.9%)	20 (9.5%)	102 (48.6%)	210 (100%)
100 season kicks	844 (40.2%)	236 (11.2%)	1020 (48.6%)	2100 (100%)
Percentage average of the two methods	41.0%	10.4%	48.6%	100%

We took the average percentage in both methods as the perceived distribution. We can see that the perceived distribution is similar under both elicitation methods. If we assume that kickers indeed kick according to these probabilities, and that goalkeepers jump according to these probabilities because they use probability matching, we obtain the predictions about the joint distribution of kicks and jumps presented in Table 11:

Table 11: Predicted Distribution if Goalkeepers Use Probability Matching (Using Perceived Distribution of Kicks)

		Goalkeeper			
		Left	Center	Right	Total
Kicker	Left	48.19	12.19	57.02	117.40
	Center	12.19	3.08	14.42	29.69
	Right	57.02	14.42	67.47	138.91
	Total	117.40	29.69	138.91	286.00

The AD value here is 116.2 and the SD value is 1961.0.

To have a convenient summary of the results in the various methods employed, Table 12 presents a summary of the AD and SD values.

Table 12: AD and SD Values in the Various Methods

Method	**AD value**	**SD value**
Maximal possible mistakes	570.0	93898.0
Kicker plays first	326.0	15614.0
Goalkeeper plays first	390.0	26332.0
Simultaneous-play methods:		
MSNE	75.2	817.0
Both play each strategy with equal probability	156.2	3585.6
Kickers use probability matching (based on the sample marginal distribution of jumps)	164.0	4361.7
Goalkeepers use probability matching (based on the sample marginal distribution of kicks)	130.2	2597.4
Both goalkeepers and kickers use probability matching	185.0	5856.6
Goalkeepers use probability matching (based on the perceived marginal distribution of kicks)	116.2	1961.0

We can see that the MSNE yields the best predictions, even though most of the other methods use information about the sample marginal distribution of kicks or jumps, which the MSNE does not use, implying that the MSNE predicts better despite using less information.

4. Conclusion

Analyzing the dataset and using several methods to predict the joint distribution of kicks and jumps suggest that the MSNE predictions are the closest to the actual sample data, even though some other methods use information on the marginal distribution of kicks or jumps whereas the MSNE does not. It should also be noted that the MSNE is not trivial, in the sense that payoffs are such that the probabilities with which each player plays each strategy are not equal (as opposed to some simple games such as matching pennies with equal payoffs, for example). Because the MSNE concept is so widely used, finding additional natural contexts in which the relevant conditions are satisfied (payoffs are common knowledge, the players have significant incentives to play correctly, the game is simple enough to analyze, its Nash equilibrium is in mixed strategies, and players' actions can be observed) and examining whether actual outcomes are close to the predictions of the MSNE is a worthwhile and important direction for future research.

Acknowledgements

We thank Ilana Ritov, an anonymous referee, and participants in a seminar given at Ben-Gurion University of the Negev and in LABSI 2008 International Conference on Strategic Decision Making in Politics and Economics in Salerno for helpful comments. We are grateful to Yael Keidar-Levin and Galit Schein for excellent research assistance. Financial support from the Phillipe Monaster Center for Economic Research is gratefully acknowledged.

References

Bar-Eli, Michael, Ofer H. Azar, Ilana Ritov, Yael Keidar-Levin and Galit Schein, 2007. Action bias among elite soccer goalkeepers: The case of penalty kicks. Journal of Economic Psychology 28(5): 606-621.

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

Budescu, David V. and Amnon Rapoport, 1994. Subjective randomization in one- and two-person games. Journal of Behavioral Decision Making 7(4): 261-278

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

Erev, Ido and Alvin E. Roth, 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. American Economic Review 88(4): 848-881.

Forrest, David and Robert Simmons, 2008. Sentiment in the betting market on Spanish football. Applied Economics 40(1): 119-126.

Hagn, Florian and Wolfgang Maennig, forthcoming. Large sport events and unemployment: the case of the 2006 soccer World Cup in Germany. Applied Economics.

Klein, Christian, Bernhard Zwergel and J. Henning Fock, forthcoming. Reconsidering the impact of national soccer results on the FTSE 100. Applied Economics.

McCabe, Kevin A, Arijit Mukherji and David E. Runkle, 2000. An experimental study of information and mixed strategy play in the three person matching pennies game. Economic Theory 15(2): 421-462.

Mookherjee, Dilip and Barry Sopher, 1994. Learning behavior in an experimental matching pennies game. Games and Economic Behavior 7(1): 62-91.

Mookherjee, Dilip and Barry Sopher, 1997. Learning and decision costs in experimental constant sum games. Games and Economic Behavior 19(1): 97-132.

Moschini, GianCarlo, 2004. Nash equilibrium in strictly competitive games: live play in soccer. Economics Letters 85(3): 365-371.

Nash, John, 1950. Equilibrium points in n person games. Proceedings of the National Academy of Sciences 36: 48-49.

Ochs, Jack, 1995. Games with unique, mixed strategy equilibria: an experimental study. Games and Economic Behavior 10(1): 202-217.

Osborne, Martin J. and Ariel Rubinstein, 1994. A Course in Game Theory. The MIT Press, Cambridge, MA.

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

Rapoport, Amnon and Richard B. Boebel, 1992. Mixed strategies in strictly competitive games: a further test of the minimax hypothesis. Games and Economic Behavior 4(2): 261-83.

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

Rapoport, Amnon and Wilfred Amaldoss, 2000. Mixed strategies and iterative elimination of strongly dominated strategies: an experimental investigation of states of knowledge. Journal of Economic Behavior and Organization 42(4): 483-521.

Rapoport, Amnon and Wilfred Amaldoss, 2004. Mixed-strategy play in single-stage first-price all-pay auctions with symmetric players. Journal of Economic Behavior and Organization 54(4), 585-607.

Roth, Alvin E. and Ido Erev, 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. Games and Economic Behavior 8(1): 164-212.

Scholtens, Bert and Wijtze Peenstra, forthcoming. Scoring on the stock exchange? The effect of football matches on stock market returns: an event study. Applied Economics.

Shachat, Jason M., 2002. Mixed strategy play and the minimax hypothesis. Journal of Economic Theory 104(1): 189-226.

Torgler, Benno and Sascha L. Schmidt, 2007. What shapes player performance in soccer? Empirical findings from a panel analysis. Applied Economics 39(18): 2355-2369.

Vulkan, Nir, 2000. An economist's perspective on probability matching. Journal of Economic Surveys 14(1): 101-118.

Walker, Mark and John Wooders, 2001. Minimax play at Wimbledon. American Economic Review 91(5): 1521-1538.