Journal of Economics and Behavioral Studies (ISSN: 2220-6140)
Vol. 13, No. 1, pp. 69-81, February 2021

Random Actions in Experimental Zero-Sum Games

Jung S. You
Department of Economics, California State University-East Bay, Hayward, CA, USA
jung.you@csueastbay.edu

Abstract: A mixed strategy, a strategy of unpredictable actions, is applicable to business, politics, and sports. Playing mixed strategies, however, poses a challenge, as the game theory involves calculating probabilities and executing random actions. I test i.i.d. hypotheses of the mixed strategy Nash equilibrium with the simplest experiments in which student participants play zero-sum games in multiple iterations and possibly figure out the optimal mixed strategy (equilibrium) through the games. My results confirm that most players behave differently from the Nash equilibrium prediction for the simplest 2x2 zero-sum game (matching-pennies) and 3x3 zero-sum game (e.g., the rock-paper-scissors game). The results indicate the need to further develop theoretical models that explain a non-Nash equilibrium behavior.

Keywords: *Zero-sum game, strategy, randomness, experiment, statistical tests*.

1. Introduction

A mixed strategy is a strategy in which a player randomly takes actions from a set of available actions, based on a set of calculated probabilities (Pindyck & Rubinfeld, 2014). Using a mixed strategy, the player would benefit from being unpredictable; thus, the other players would not be able to predict which action is going to be played (Bernheim & Whinston, 2013; McCain, 2014). The application of mixed strategies extends to dealing with terrorism, tax evasion, playing poker, beating the stock market, and winning a new product market against competitors.

Since O'Neill (1987) and Brown & Rosenthal (1990), there has been mixed evidence regarding the empirical validity of the Nash equilibrium in mixed strategies. Negative evidence suggests that mixing does not occur as predicted, particularly when the data is analyzed at the individual level.[1] Positive evidence suggests that professional players play mixed strategies according to equilibrium predictions in competitive sports, such as soccer, tennis, and baseball, within which there are a winner and a loser[2]. Several articles show both positive and negative results. Walker & Wooders (2001) find that the expected payoff of a pure strategy is almost the same as that of another pure strategy in each 2x2 stage game, which is consistent with the equilibrium predictions. However, they note that the players' choices exhibit serial correlation. Van Essen & Wooders (2015) show that professional players' behaviors are closer to equilibrium than novices'. Emara et al. (2017) examined strategic decisions in the National Football League and investigated whether the chosen sequence exhibits serial correlation. The authors find that the choices of professional players exhibit serial correlation. And, more recently, using data from the rock-paper-scissors games played on a Facebook application, Batzilis et al. (2019) show that players deviate from the Nash equilibrium in response to the information of their opponent's historical matches with other players[3].

In this paper, I investigate how closely the mixed strategy Nash equilibrium theory can predict individual behavior in finitely repeated zero-sum games: matching-pennies and rock-paper-scissors[4]. The unique Nash

[1] This literature includes Brown & Rosenthal (1990), O'Neill (1991), Rapoport & Boebel (1992), Mookherjee & Sopher (1994), Ochs (1995), Mookherjee & Sopher (1997), Rapoport & Amaldoss (2000), Shachat (2002), Camerer (2011), Levitt et al. (2010), and Duffy et al. (2021).
[2] This literature includes Binmore et al. (2001), Chiappori et al. (2002), Palacios-Huerta (2003), Coloma (2007), Palacios-Huerta & Volij (2008), Azar & Bar-Eli (2011), and Buzzacchi & Pedrini (2014).
[3] The analysis in Batzilis et al. (2019) focuses on the first round in each match as non-equilibrium behavior on subsequent rounds involves a player's response to the prior rounds in the match in addition to the opponent's history with other players. Their games are effectively one-shot simultaneous-move games with the opponent's history before the match.
[4] The rock-paper-scissor game has championship competitions around the world (for instance, see the World Rock Paper Scissors Association: https://www.wrpsa.com/rock-paper-scissors-world-championship/).

equilibrium in a finitely repeated zero-sum game is playing the same one-shot game Nash equilibrium in each round. Repeated two-person zero-sum games produce multiple observations of individual behavior for testing Nash equilibrium assumptions of both best responses and backward induction.

I use two-person matching-pennies and rock-paper-scissors since they are the simplest among all the games that have been tested in the literature (most literature use games with which participants are unfamiliar and in which it may not be possible for them to become proficient in the limited timeframe of experiments). The payoff in matching-pennies and rock-paper-scissors includes only two outcomes of wins and losses; the games are symmetric with respect to the mixed strategies of the row and column players and gameplay is face-to-face. Due to their simplicity in structure and equilibrium solution, players face a cognitively less demanding task than those in other experiments. One can expect that nonprofessional player quickly learn both games and might approximate the behavior in the mixed strategy Nash equilibrium.

In my experiment, 36 pairs of subjects played a matching-pennies game or the rock-paper-scissor game 20 times in succession, face to face. The players were told in advance the exact number of repetitions. The matching-pennies game has a unique equilibrium: both players randomize with probabilities .5 and .5, respectively. The rock-paper-scissor game's unique equilibrium is that both players mix according to probabilities .33, .33, and .33, respectively. As in the typical experimental setting, the game has a precisely defined set of rules, only a few strategies are available, outcomes are decided immediately after strategies are chosen, and all relevant information is observable.

With the data collected from the experiment, I test two hypotheses that equilibrium theory of best responses and backward induction yields the behavior of players in repeated zero-sum games. The first hypothesis is that at every stage, players choose pure strategies with equal mixture probabilities (according to the equilibrium strategy of the symmetric one-shot game). The second hypothesis is that equilibrium strategies in finitely repeated zero-sum games are independent of the time lags between the stages of the game.

The results of the tests do not support equilibrium play for most participants. The first null hypothesis that the mixed probabilities for individual players are identical across pure strategies in each repetition of symmetric zero-sum games is rejected at a 10% significance level for all 72 participants except four players under the matching-pennies game. The second hypothesis that players generate serially independent sequences in repeated games is rejected at a 10% significance level for more than half of a total of 72 participants. The hypothesis rejection occurs for 83% of the players under the rock-paper-scissor games whereas it is rejected for 31% of the players under the matching-pennies games.

The results in this paper support the need for developing theoretical models of an individual's non-equilibrium plays. The results also support my observation and feedback from students and teachers that mixed strategies are difficult to understand and to implement. The difficulty of using mixed strategies resides in calculating mixture probabilities and using a random device according to chosen probabilities. In fact, people are known to have difficulty generating random numbers.[5]

The remainder of the paper is organized as follows. Section 2 describes the structure and setting of the play. Section 3 is devoted to the empirical analysis. Finally, Section 4 concludes.

2. Zero-Sum Game Experiments

A total of 72 students with various majors from a public university in California participated in the experiments from fall 2018 to fall 2019.

[5] Research has shown that people have difficulty with detecting and producing random sequences of the sort required for the execution of mixed strategies (Bar-Hillel & Wagenaar, 1991; Oskarsson et al., 2009; Rabin, 2002; Rapoport & Budescu, 1992; Willem A. Wagenaar, 1972).

The Matching-Pennies Games: At the beginning of fall 2018, a group of 36 students played the simplest zero-sum game the matching-pennies games.[6]

The matching-pennies game involves two players. All players were informed in advance that the game would be played 20 times against the same opponent. One player in each pair is referred to as a row player and the other as a column player. Each player holds a penny and displays either head (H) or tails (T) simultaneously in each round. A row player gains one point, and a column player loses one point if the coins show the same side, and a row player loses one point and a column player gains one point if the coins show different sides, in each round. Table 1 shows the bimatrix form of the game where the left number in each cell is the payoff of a row player. Participants were not shown the bimatrix form. Since each player wins 1 or loses 1 depending on the opponent for the same side, the best strategy, in theory, is to play a mixed strategy in which each player chooses H or T with equal chance, that is, unbiased coin toss.

Table 1: Matching-Pennies

		Column Player	
		H	T
Row Player	H	1, -1	-1, 1
	T	-1, 1	1, -1

At the time when students played the zero-sum game, they had not learned the concept of mixed strategies. On the instructional handout, students had to write their strategies to win the game and had to calculate payoffs. The following box provides the class handout for the game (the rounds after the first round are omitted due to repetition in the table):

Matching-Pennies

The instructor has you, group, into pairs. Decide who is a row player or column player.
Row Player: Column Player:
Each of you shows a coin simultaneously. Row Player's payoff is 1 (or -1) and Column Player's is -1 (or +1), when the two coins show the same side (or different sides), respectively. For each round, write down your strategy in the "Your action" column to play Head (H) or Tail (T) before you and your opponent take simultaneous actions. At the end of each round, write down your payoff for that round and compute your cumulative and average payoffs in the following table.

	Your action (H or T)	Opponent's action	Your payoff	Your cumulative payoff	Your average payoff
Round 1					

To incentivize the students to play the game seriously, the author announced that each student's cumulative payoffs after 20 rounds would be proportionally converted into participation credits that they needed to collect for their final grades.

The Rock-Paper-Scissors Games: At the beginning of fall 2019, another group of 36 students played the well-known zero-sum game RPS.[7] The RPS game also involves two players. All players were informed in advance that the game would be played 20 times against the same opponent. In RPS, each player simultaneously forms one of three shapes, "rock" (R, a closed fist), "paper" (P, a flat hand), or "scissors" (S, a fist with the index finger and middle finger extended, forming a V), with an outstretched hand. A player who plays R will beat another player who has chosen S but will lose to one who has played P. A play of P will lose to a play of S. If both players choose the same shape, the game is tied.

[6] You (2019) analyzed the same experiments of the matching-pennies games.
[7] Instances of use of RPS in real-life scenarios include an American court case (Avista Management v. Wausau Underwriters in 2016), an auction house selection between Christie's and Sotheby's for the collection of Impressionist paintings in 2005, and numerous RPS games around the world including World Rock Paper Scissors Society that started in 2002.

Table 2 shows the bimatrix form of the game where the left number in each cell is the payoff of the row player. Participants were not shown the bimatrix form. The best strategy, in theory, is to play a mixed strategy in which each player chooses R, P, or S with equal chance, that is, an unbiased dice toss where number 1 or 2 indicates a play of R, 3 or 4 a play of P, and 5 or 6 a play of S.

Table 2: Rock-Paper-Scissors

		Column Player		
		R	P	S
	R	0, 0	-1, 1	1, -1
Row Player	P	1, -1	0, 0	-1, 1
	S	-1, 1	1, -1	0, 0

Students who participated in the experiment had not learned the concept of mixed strategies. On the instructional handout, students had to write their strategies to win the game and had to calculate payoffs. The following box provides the class handout for the game (The rounds after the first round are omitted due to repetition in the table):

Rock-Paper-Scissors

The instructor has you, group, into pairs. Decide who is a row player or column player.
Row Player: Column Player:
You play an RPS zero-sum game. Each of you shows your hand sign simultaneously. Each player simultaneously forms one of three shapes with an outstretched hand. These shapes are "rock" (a closed fist), "paper" (a flat hand), and "scissors" (a fist with the index finger and middle finger extended, forming a V). You know the rule: rock crushes scissors, paper covers rock, scissors cuts paper. The winner scores 1 and the loser scores -1. If the game is tied, each earns 0. For each round, write down your strategy in the "Your action" column to play Rock (R), Paper (P) or Scissors (S) before you and your opponent take simultaneous actions. At the end of each round, write down your payoff for that round and compute your cumulative and average payoffs in the following table.

	Your action (R, P, or S)	Opponent's action	Your payoff	Your cumulative payoff	Your average payoff
Round 1					

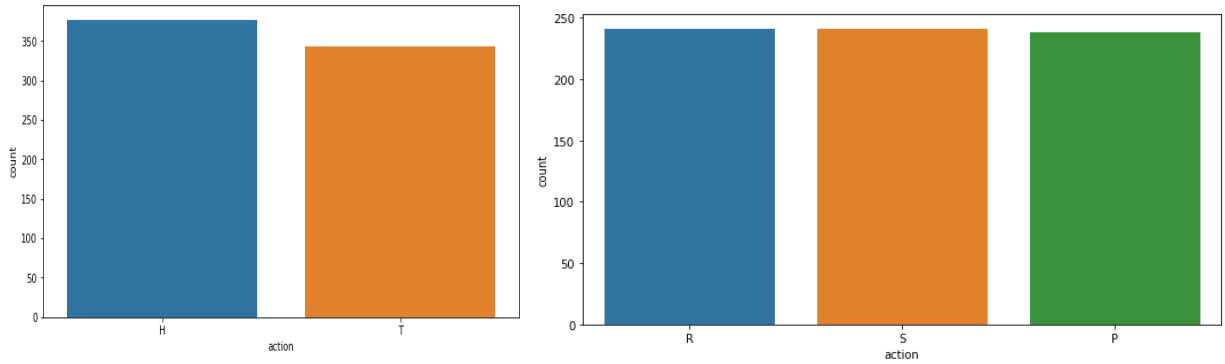
To incentivize the students to play the game seriously, the author announced that each student's cumulative payoffs after 20 rounds would be proportionally converted into participation credits that they needed to collect for their final grades.

3. Empirical Analysis

The marginal distributions of the aggregate data from both games seemingly exhibit uniformly distributed choices. For the matching-pennies game, my data consist of records of actions and payoffs that 36 students in 18 pairs submitted after a total of 20 rounds of the game. I cross-checked the actions and payoff calculations the students reported and did not find any mistakes in them. Heads were chosen 377 times, and tails were chosen 343 times out of a total of 720 times. The frequency of heads was 0.52 and that of tails was 0.48, close to .5 and .5. Figure 1(a) is the aggregate histogram of heads and tails after 20 rounds of the game with 36 students, and its appearance is close to that of a uniform distribution.

Figure 1

(a) "Matching-Pennies" Class Histogram (b) "Rock-Paper-Scissors" Class Histogram



For the RPS game, my data consist of records of actions and payoffs that another 36 students in 18 pairs submitted after a total of 20 rounds of the game. I cross-checked the actions and payoff calculations the students reported and did not find any mistakes in them. In total, R appeared 241 times, P appeared 238 times, and S appeared 241 times; thus, the frequency of each shape makes .33, .33, and .33, respectively. Figure 1(b) is the aggregate histogram of R, P, and S after 20 rounds of the game with 36 students. Its appearance is close to that of a uniform distribution.

Game theory models individuals' rational behavior. Since aggregate data average out individual differences, I ask whether individual players' observed choices match the theoretical predictions.[8] The unique Nash equilibrium for both games predicts that in each round (i) players choose each action with equal probability and (ii) their choices are independent of their previous actions and their opponents' choices.

Individual Tests of Equal Mixture Probabilities: The tests of the null hypothesis that the mixture probabilities for individual players are identical across pure strategies in each repetition can be implemented with the Kolmogorov—Smirnov test of discrete uniform distributions. For the matching-pennies game, I apply a one-sample Kolmogorov—Smirnov test for each of a total of 36 individual players where I record tails as 0 and heads as 1.[9] The discrete Kolmogorov—Smirnov goodness-of-fit test (KS test) is an alternative to the Chi-square test, which does not achieve high statistical power for small sample sizes for discrete null distributions (Horn, 1977; Slakter, 1965). For binomial distributions, the p-values are known to be exact for the KS test (Arnold & Emerson, 2011). Using the dg of R package for discrete null distribution, I estimate the p-value via a Monte Carlo simulation with 10,000 replicates. Table 3 shows the observed frequency or mixture of choices, the Kolmogorov—Smirnov test statistic, and its p-value for Player 1 to Player 36.

Table 3: Matching-Pennies Game

Tests for equality of mixture probabilities

		Mixture		**Kolmogorov—Smirnov**		
Player	**#Obs**	**H**	**T**	**Statistic**	**P-Value**	
1	20	0.6	0.4	0.967	0.000	***
2	20	0.15	0.85	0.500	0.000	***
3	20	0.7	0.3	0.350	0.003	***
4	20	0.5	0.5	0.450	0.000	***
5	20	0.55	0.45	0.300	0.012	**
6	20	0.4	0.6	0.300	0.011	**
7	20	0.4	0.6	0.300	0.012	**

[8] You (2019) performs the McNemar test on the joint distribution of paired plays. The test is based on aggregate averaged behavior that does not explain individual differences.

[9] The one-sample Kolmogorov—Smirnov test is a nonparametric test of the equality of (continuous or discontinuous) one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution.

8	20	0.6	0.4	0.350	0.002	***
9	20	0.5	0.5	0.350	0.003	***
10	20	0.45	0.55	0.400	0.000	***
11	20	0.65	0.35	0.200	0.120	
12	20	0.55	0.45	0.350	0.003	***
13	20	0.7	0.3	0.300	0.013	**
14	20	0.45	0.55	0.300	0.010	**
15	20	0.55	0.45	0.300	0.012	**
16	20	0.5	0.5	0.200	0.117	
17	20	0.6	0.4	0.250	0.041	*
18	20	0.5	0.5	0.350	0.003	***
19	20	0.4	0.6	0.300	0.012	**
20	20	0.7	0.3	0.350	0.002	***
21	20	0.5	0.5	0.500	0.000	***
22	20	0.55	0.45	0.200	0.117	
23	20	0.65	0.35	0.350	0.003	***
24	20	0.45	0.55	0.200	0.040	*
25	20	0.45	0.55	0.333	0.000	***
26	20	0.55	0.45	0.133	0.198	
27	20	0.25	0.75	0.433	0.000	***
28	20	0.75	0.25	0.300	0.001	***
29	20	0.45	0.55	0.333	0.000	***
30	20	0.5	0.5	0.300	0.001	***
31	20	0.65	0.35	0.333	0.000	***
32	20	0.35	0.65	0.567	0.000	***
33	20	0.65	0.35	0.567	0.000	***
34	20	0.5	0.5	0.300	0.001	***
35	20	0.55	0.45	0.267	0.005	***
36	20	0.6	0.4	0.200	0.042	*

Note: *** Indicates rejected at 1% level, ** indicates rejected at 5% level, and * indicates rejected at 10% level.

The results in Table 3 show that the null hypothesis that mixing probabilities are identical across strategies is rejected for most players. Of the 36 players in the sample, the hypothesis is rejected for 22 players at the 1% significance level, 29 players at the 5% significance level, and 32 players at the 10% significance level. The hypothesis cannot be rejected for the remaining four players (Players 11, 16, 22, and 26) since the corresponding p-values range from 0.117 to 0.198. Given that the power of the KS test must be low for the small sample size for each player, the hypothesis could have been rejected with more observations.

Table 4: Rock-Paper-Scissors Game

Tests for equality of mixture probabilities

		Mixture			**Kolmogorov—Smirnov**		
Player	**#Obs**	**R**	**P**	**S**	**Statistic**	**P-Value**	
37	20	0.3	0.2	0.5	0.333	0.003	***
38	20	0.45	0.2	0.35	0.333	0.003	***
39	20	0.05	0.65	0.3	0.367	0.002	***
40	20	0.45	0.25	0.3	0.367	0.001	***
41	20	0.4	0.3	0.3	0.367	0.002	***
42	20	0.35	0.35	0.3	0.367	0.002	***
43	20	0.2	0.45	0.35	0.333	0.002	***
44	20	0.25	0.4	0.35	0.333	0.002	***
45	20	0.3	0.3	0.4	0.333	0.002	***

46	20	0.3	0.5	0.2	0.467	0.000	***
47	20	0.4	0.25	0.35	0.333	0.003	***
48	20	0.4	0.3	0.3	0.367	0.003	***
49	20	0.2	0.5	0.3	0.367	0.001	***
50	20	0.5	0.2	0.3	0.367	0.002	***
51	20	0.7	0.1	0.2	0.467	0.000	***
52	20	0.45	0.3	0.25	0.417	0.000	***
53	20	0.3	0.35	0.35	0.333	0.002	***
54	20	0.4	0.3	0.3	0.367	0.001	***
55	20	0.3	0.35	0.35	0.333	0.002	***
56	20	0.25	0.25	0.5	0.333	0.004	***
57	20	0.3	0.55	0.15	0.517	0.000	***
58	20	0.25	0.4	0.35	0.333	0.002	***
59	20	0.35	0.45	0.2	0.467	0.000	***
60	20	0.5	0.15	0.35	0.333	0.003	***
61	20	0.35	0.45	0.2	0.467	0.000	***
62	20	0.25	0.25	0.5	0.333	0.002	***
63	20	0.3	0.3	0.4	0.333	0.002	***
64	20	0.35	0.25	0.4	0.333	0.002	***
65	20	0.3	0.35	0.35	0.333	0.002	***
66	20	0.35	0.35	0.3	0.367	0.002	***
67	20	0.2	0.25	0.55	0.333	0.002	***
68	20	0.2	0.25	0.55	0.333	0.003	***
69	20	0.4	0.35	0.25	0.417	0.000	***
70	20	0.35	0.35	0.3	0.367	0.002	***
71	20	0.35	0.45	0.2	0.467	0.000	***
72	20	0.3	0.25	0.45	0.333	0.002	***

Note: *** Indicates rejected at 1% level, ** indicates rejected at 5% level, and * indicates rejected at 10% level.

For the RPS game, the unique Nash equilibrium is for every player to choose R, P, and S with an equal probability (i.e., .33), in each round. To check whether this theoretically optimal strategy was implemented, I applied the Kolmogorov—Smirnov test for each player where I record R as 0, P as 1, and S as 2. Using the dgof R package for discrete null distribution, I estimate the p-value via a Monte Carlo simulation with 10,000 replicates. Table 4 shows the observed frequency of choices, the KS test statistic and its p-value for Player 37 to Player 72.

The results in Table 4 show that the null hypothesis is rejected for all 36 players at the 1% significance level. These estimates suggest that at the individual level, the hypothesis that mixing probabilities are identical across strategies is rejected for all players at a conventional significance level.

Individual Tests of Serial Independence

In this section, I ask whether players' observed choices can be modeled as i.i.d. drawings from pair-specific stationary distributions.

Individual Tests of Serial Independence for the Matching-Pennies Game: For the matching-pennies game, I ask whether players' observed choices can be modeled as i.i.d. drawings from pair-specific stationary logit distributions. I investigate the possibility that both one's own and one's opponent's past plays are used to condition current plays. I here test whether this condition holds, and when it does not hold, I identify the sources of the failure.

To confirm that past choices have no role in determining current choices, I estimate logistic regressions for each player, applying the analysis of Palacios-Huerta (2003). My dependent variable takes a value of 1 if the play heads and a 0 otherwise. The independent variables in these equations were: first lagged indicators for both players' past choices and an indicator for the opponent's current choice. The latter is included to allow for the possibility that a player might be able to "read the face" of his or her opponent. I have experimented

with the inclusion of second lags and the lagged interaction term of the two players' choice indicators in the equations underlying Table 5. However, I have found these terms to be statistically unimportant in explaining current choices for all players.

I then performed the likelihood-ratio tests of significance for the joint influence of lagged own choices, lagged opponent's choices, and contemporaneous opponent's choices. The results of the five hypothesis tests are summarized in Table 5. The first test measures the joint significance of all explanatory variables in accounting for a player's choice. According to the mixed strategy Nash equilibrium model, all the explanatory variables should be extraneous, and so we should be unlikely to find many cases in which these variables appear to be important in explaining players' choices. Only for three of the 18 pairs, at least one player's behavior is significantly determined at the 10% significance level by the set of explanatory variables included in my equations.

Table 5: Matching-Pennies Game

Results of significance tests from a logit equation for the choice of 'head'

Estimating Equation: $\Pr(C) = \frac{\exp[a_0 + a_1 \text{ lag}(C) + b_0 C^* + b_1 \text{lag}(C^*)]}{1+\exp[a_0 + a_1 \text{ lag}(C) + b_0 C^* + b_1 \text{lag}(C^*)]}$

		Players whose behavior allows rejection of the null hypothesis at the:			
Null hypothesis:		0.001 level	0.01 level	0.05 level	0.10 level
1 $a_1 = b_0 = b_1 = 0$	Row:			11,17	11,17
	Column:		18	18	12,18,32
2 $b_0 = b_1 = 0$	Row:			11,17	11,17,27
	Column:		18	12,18	12,18,30
3 $b_1 = 0$	Row:			7	7
	Column:			18,30	18,30
4 $b_0 = 0$	Row:		11,17	11,17	11,17,27
	Column:		12,18	12,18	12,18,28
5 $a_1 = 0$	Row:			17	17,31
	Column:			20,32	20,32

Notes: Tails is a pivot outcome.
C and C* denote the choice of a player and his or her opponent, respectively.
The terms 'lag' refers to the strategies previously followed in the ordered sequence of games.
Rejections are based on likelihood-ratio tests.

The observable correlations found in players' choices could have been exploited by their opponents. I, therefore, look for evidence of the influence of opponents' choices. The second test summarized in Table 5 measures the significance of terms involving the opponent's current and lagged choices in determining a player's current choice. If players intended to play the mixed strategy Nash equilibrium and if players believed their opponents to be playing the mixed strategy Nash equilibrium, the various terms involving the opponent's current and lagged plays should not influence that player's choices. For four of the 18 pairs, at least one player's behavior is significantly influenced at the 10% significance level by the set of explanatory variables included in my equations.

The third test reported in Table 5 concerns the significance of the linear terms involving opponents' lagged play. If players attempted to predict their opponents' current choices partly based on the opponents' past choices, one would expect this term to influence observed choices. For three of the 18 pairs, at least one player's behavior is significantly determined at the 10% significance level by the set of explanatory variables included in my equations. If we cannot reject the unimportance of opponents' plays in determining players' choices, then we may accept the notion that the players themselves believed their opponents to be playing the mixed strategy Nash equilibrium. As indicated in Table 5, the data seem consistent with this notion for many players.

The fourth test reported in Table 5 concerns the ability of one player to discern the current choice of his or her opponent, even after controlling for the influence of past choices. My results indicate that "face reading" may have occurred for three pairs at the 10% significance level.

The fifth test focuses on the explanatory importance of a player's own past choices in determining his or her current choice. For three of the 18 pairs, at least one player's behavior is significantly determined at the 10% significance level by the set of explanatory variables included in my equations.

The results in Table 5 for the finitely repeated matching-pennies game might imply that nonprofessional players can make serially independent choices for each stage game as if they apply backward induction or try to play unpredictable moves. The reason why most players failed to play heads and tails with equal probabilities might be in the difficulty of generating random actions.[10]

Individual Tests of Serial Independence for the RPS Game: For the RPS game, I ask whether players' observed choices can be modeled as i.i.d. drawings from pair-specific stationary multinomial logit distributions. I estimate multinomial logistic regressions for each player, applying the analysis of Brown & Rosenthal (1990). The dependent variable is a trichotomous indicator of the choice of gesture from R, P, and S. My dependent variable takes a value of 0 if the play is R, 1 if P, and 2 otherwise. The independent variables are a first lagged indicator for both players' past choices and an indicator for the opponent's current choices. I then performed the likelihood-ratio tests of significance for the joint influence of lagged own choices, lagged opponent's choices, and contemporaneous opponent's choices. The results of the five hypothesis tests are summarized in Table 6.

Table 6: Rock-Paper-Scissors Game

Results of significance tests from multinomial logit equation for the choice of handshape

Estimating Equation:

$$\Pr(C_k) = \frac{\exp\left[a_{0,k}+a_{1,k}\mathrm{lag}(C_k)+b_{0,k}C^*+b_{1,k}\,\mathrm{lag}(C^*)\right]}{\left[1+\exp\left[a_{0,k}+a_{1,k}\mathrm{lag}(C_k)+b_{0,k}\,C^*+b_{1,k}\,\mathrm{lag}(C^*)\right]+\exp\left[a_{0,k'}+a_{1,k'}\mathrm{lag}(C_{k'})+b_{0,k'}C^*+b_{1,k'}\mathrm{lag}(C^*)\right]\right]} \text{ for } k, k'=1, 2, k\neq k'$$

		Players whose behavior allows rejection of the null hypothesis at the:		
Null hypothesis:			0.001 level	0.01 level
1	$a_1 = b_0 = b_1 = 0$	Row:	47	37,41,43,47,61,63,65,67,69
		Column:		38,42,52
2	$b_0 = b_1 = 0$	Row:	37,41	37,41,43,53,63,67,69
		Column:	42	38,42,50
3	$b_1 = 0$	Row:		41,43
		Column:		38,42,50
4	$b_0 = 0$	Row:	37,41	37,41,69
		Column:	42	38,42,48,50
5	$a_1 = 0$	Row:	47	37,41,47,61,63,71
		Column:		58
Null hypothesis:			0.05 level	
1	$a_1 = b_0 = b_1 = 0$	Row:	37,41,43,47,49,53,59,61,63,65,67,69,71	
		Column:	38,42,44,48,50,52,58,66,68,70,72	
2	$b_0 = b_1 = 0$	Row:	37,41,43,47,49,53,61,63,67,69	
		Column:	38,42,48,50,52,58,64,68,70,72	
3	$b_1 = 0$	Row:	41,43,53,61,63,67	

[10] Some literature find that people have difficulty with producing independent, random sequences. For instance, see W. A. Wagenaar, (1970), W. A. Wagenaar (1972).

		Column:	38,42,48,50,58,62,64,72
4	$b_0 = 0$	Row:	37,41,49,65,67,69,71
		Column:	38,42,48,50,52,64,68,70,72
5	$a_1 = 0$	Row:	37,41,43,47,49,53,59,61,63,65,71
		Column:	42,44,50,52,56,58
Null hypothesis:			0.10 level
1	$a_1 = b_0 = b_1 = 0$	Row:	37,41,43,47,49,51,53,59,61,63,65,67,69,71
		Column:	38,42,44,48,50,52,54,58,64,66,68,70,72
2	$b_0 = b_1 = 0$	Row:	37,41,43,47,49,51,53,59,61,63,65,67,69,71
		Column:	38,42,44,48,50,52,58,62,64,66,68,70,72
3	$b_1 = 0$	Row:	37,41,43,53,59,61,63,67,69
		Column:	38,42,44,48,50,58,62,64,72
4	$b_0 = 0$	Row:	37,41,49,51,59,65,67,69,71
		Column:	38,42,48,50,52,60,62,64,66,68,70,72
5	$a_1 = 0$	Row:	37,41,43,47,49,53,59,61,63,65,71
		Column:	42,44,50,52,56,58

Notes: R is a pivot outcome, $C_1 = P$, and $C_2 = S$.
C and C* denote the choice of a player and his or her opponent, respectively.
The terms 'lag' refers to the strategies previously followed in the ordered sequence of games.
Rejections are based on likelihood-ratio tests

The first test measures the joint significance of all explanatory variables in accounting for a player's choice. For 15 of the 18 pairs, at least one player's behavior is significantly determined at the 10% significance level by the set of explanatory variables included in my equations. The statistical significance of these variables in explaining players' choices is the rule rather than the exception. I take this finding as strong evidence against the mixed strategy Nash equilibrium model.

The second test summarized in Table 6 measures the significance of terms involving the opponent's current and lagged choices in determining a player's current choice. If players intended to play the mixed strategy Nash equilibrium and if players believed their opponents to be playing the mixed strategy Nash equilibrium, the various terms involving the opponent's current and lagged plays should not influence that player's choices. However, for 15 of the 18 pairs, at least one player's behavior is significantly determined at the 10% significance level by the set of explanatory variables included in my equations.

The third test reported in Table 6 concerns the significance of the linear terms involving opponents' lagged play. The results of the third test are that for nine-row players and nine-column players, the set of terms involving opponents' lagged plays are statistically significant at the .10 level in determining current choices. In other words, for 13 of the 18 pairs, at least one player's behavior is significantly influenced by this term. The fourth test reported in Table 6 concerns the ability of one player to discern the current choice of his or her opponent, even after controlling for the influence of past choices. For 12 of the 18 pairs, at least one player's behavior is significantly influenced by this “face reading” term. Lastly, the results of the fifth test indicate that a substantial number of players made choices that were significantly related to their own previous choices. For 11 row players and six-column players, one can reject at the .10 level the null hypothesis that one’s own past choices are uncorrelated with current choices. Only for two-row players (Players 40 and 46) and four column players (Players 39, 45, 55, and 57), any of the five null hypotheses cannot be rejected at the .10 level. In other words, 83% of the players behaved differently from the theoretical prediction of serial independence.

The findings in Table 6 indicate that the choices of at least three-quarters of players are related to their own previous choices and opponents’ current and previous choices. These rejections occur despite the low power associated with pair-by-pair tests and even though some forms of behavior not in mixed strategy Nash

equilibrium can generate a mixture. The most informative rejections come from the interdependent choices shown by so many player pairs. The results show little support for the mixed strategy model, even when this paper assumes stationarity in the process of generating players' choices.

4. Conclusion

My results confirm that most players behave differently from the Nash equilibrium prediction for the simplest matching-pennies and rock-paper-scissors games. The individual KS tests show that the first null hypothesis that the mixed probabilities for individual players are identical across pure strategies in each repetition of symmetric zero-sum games is rejected at a 10% significance level for all 72 participants except four players under the matching-pennies game. In addition, the individual likelihood ratio tests show that the second hypothesis that players generate serially independent sequences in repeated games is rejected at a 10% significance level for more than half of a total of 72 participants. The hypothesis rejection occurs for 83% of the players under the rock-paper-scissor games whereas it is rejected for 31% of the players under the matching-pennies games. The results of the tests do not support Nash equilibrium play for most participants.

The effects of the computational difficulties of mixed strategies could explain the results of deviations from Nash equilibrium prediction, especially for the finitely repeated RPS game.[11] Another possible explanation is that players have difficulties concealing hand gestures to their opponents even though they try to play random actions. Once a player reads the oncoming shape of the opponent's hand that is seemingly a non-equilibrium play. The player's best response is to shift from the mixed strategy Nash equilibrium and to play the corresponding move.[12] In addition to this, as people have difficulty producing independent, random sequences, serial independent plays might not only be irrational but also impossible for human subjects.[13] The results in this paper could be improved by greater sample sizes and higher incentives for players' competitive behavior. Nevertheless, given that the games this paper investigates are the simplest among all two-player games studied in the literature, the results support evidence against the validity of mixed strategy Nash equilibrium to describe most human behavior.

Studying evolutionary game theory to model an individual's non-equilibrium plays is a potentially fruitful direction for future research. Recent studies show that evolutionary game theory outperforms Nash equilibrium in predicting average mixed strategies in asynchronous RPS-like games (Cason et al., 2014; Hoffman et al., 2015).[14]

References

Arnold, T. B. & Emerson, J. W. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. *R Journal*, 3(2).

Azar, O. H. & Bar-Eli, M. (2011). Do soccer players play the mixed-strategy Nash equilibrium? *Applied Economics*, 43(25), 3591–3601.

Bar-Hillel, M. & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics*, 12(4), 428–454.

Batzilis, D., Jaffe, S., Levitt, S., List, J. A. & Picel, J. (2019). Behavior in Strategic Settings: Evidence from a Million Rock-Paper-Scissors Games. *Games*, 10(2), 18.

Benaïm, M., Hofbauer, J. & Hopkins, E. (2009). Learning in games with unstable equilibria. *Journal of Economic Theory*, 144(4), 1694–1709.

[11] For an example of computational difficulties, see Halpern & Pass (2015).

[12] A robot can beat the best human players in the RPS games, by using a high-speed camera and electronic reflexes to identify the oncoming shape of the opponent's hand and play the corresponding move to beat it.

[13] Randomness is easier to disprove than to prove since for disproving randomness, it is sufficient to show one type of systematic trend (W. A. Wagenaar, 1970; W. A. Wagenaar, 1972). This applies to testing on mixed strategy equilibrium.

[14] Cason et al. (2014) show that Time Average of the Shapley Polygon (TASP) outperforms Nash equilibrium in predicting average mixed strategies in asynchronous RPS-like games. Benaïm et al. (2009) propose TASP as an alternative model of non-equilibrium behavior in games whose Nash equilibria are mixed strategies.

Bernheim, D. & Whinston, M. (2013). Microeconomics.

Binmore, K., Swierzbinski, J. & Proulx, C. (2001). Does minimax work? An experimental study. Economic Journal, 445–464.

Brown, J. N. & Rosenthal, R. W. (1990). Testing the minimax hypothesis: A re-examination of O'Neill's game experiment. Econometrica: Journal of the Econometric Society, 1065–1081.

Buzzacchi, L. & Pedrini, S. (2014). Does player specialization predict player actions? Evidence from penalty kicks at FIFA World Cup and UEFA Euro Cup. *Applied Economics*, 46(10), 1067–1080.

Camerer, C. F. (2011). Behavioral game theory: Experiments in strategic interaction. Princeton university press.

Cason, T. N., Friedman, D. & Hopkins, E. (2014). Cycles and instability in a rock–paper–scissors population game: A continuous-time experiment. *Review of Economic Studies*, 81(1), 112–136.

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

Coloma, G. (2007). Penalty kicks in soccer: An alternative methodology for testing mixed-strategy equilibria. *Journal of Sports Economics*, 8(5), 530–545.

Duffy, S., Naddeo, J., Owens, D. M. & Smith, J. (2021). Cognitive load and mixed strategies: On brains and minimax. Available at SSRN 3770824.

Emara, N., Owens, D., Smith, J. & Wilmer, L. (2017). Serial correlation in National Football League plays calling and its effects on outcomes. *Journal of Behavioral and Experimental Economics*, 69, 125–132.

Halpern, J. Y. & Pass, R. (2015). Algorithmic rationality: Game theory with costly computation. *Journal of Economic Theory*, 156, 246–268.

Hoffman, M., Suetens, S., Gneezy, U. & Nowak, M. A. (2015). An experimental investigation of evolutionary dynamics in the Rock-Paper-Scissors game. *Scientific Reports*, 5(1), 1–7.

Horn, S. D. (1977). Goodness-of-fit tests for discrete data: A review and an application to a health impairment scale. Biometrics, 237–247.

Levitt, S. D., List, J. A. & Reiley, D. H. (2010). What happens in the field stays in the field: Exploring whether professionals play minimax in laboratory experiments. *Econometrica*, 78(4), 1413–1434.

McCain, R. A. (2014). Game theory: A non-technical introduction to the analysis of strategy. World Scientific Publishing Company.

Mookherjee, D. & Sopher, B. (1994). Learning behavior in an experimental matching pennies game. *Games and Economic Behavior*, 7(1), 62–91.

Mookherjee, D. & Sopher, B. (1997). Learning and decision costs in experimental constant sum games. *Games and Economic Behavior*, 19(1), 97–132.

Ochs, J. (1995). Games with unique, mixed strategy equilibria: An experimental study. *Games and Economic Behavior*, 10(1), 202–217.

O'Neill, B. (1987). Nonmetric test of the minimax theory of two-person zero-sum games. *Proceedings of the National Academy of Sciences*, 84(7), 2106–2109.

O'Neill, B. (1991). Comments on Brown and Rosenthal's reexamination. Econometrica: Journal of the Econometric Society, 503–507.

Oskarsson, A. T., Van Boven, L., McClelland, G. H. & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin*, 135(2), 262.

Palacios-Huerta, I. (2003). Professionals play minimax. *The Review of Economic Studies*, 70(2), 395–415.

Palacios-Huerta, I. & Volij, O. (2008). Experientia docet: Professionals play minimax in laboratory experiments. *Econometrica*, 76(1), 71–115.

Pindyck, R. S. & Rubinfeld, D. L. (2014). Microeconomics. Pearson Education.

Rabin, M. (2002). Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, 117(3), 775–816.

Rapoport, A. & Amaldoss, W. (2000). Mixed strategies and iterative elimination of strongly dominated strategies: An experimental investigation of states of knowledge. *Journal of Economic Behavior & Organization*, 42(4), 483–521.

Rapoport, A. & Boebel, R. B. (1992). Mixed strategies in strictly competitive games: A further test of the minimax hypothesis. *Games and Economic Behavior*, 4(2), 261–283.

Rapoport, A. & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General*, 121(3), 352.

Shachat, J. M. (2002). Mixed strategy play and the minimax hypothesis. *Journal of Economic Theory*, 104(1), 189–226.

Slakter, M. J. (1965). A comparison of the Pearson chi-square and Kolmogorov goodness-of-fit tests with respect to validity. *Journal of the American Statistical Association*, 60(311), 854–858.

Van Essen, M. & Wooders, J. (2015). Blind stealing: Experience and expertise in a mixed-strategy poker experiment. *Games and Economic Behavior*, 91, 186–206.

Wagenaar, W. A. (1970). Appreciation of conditional probabilities in binary sequences. *Acta Psychologica*, 34, 348–356.

Wagenaar, Willem A. (1972). Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin*, 77(1), 65.

Walker, M. & Wooders, J. (2001). Minimax play at Wimbledon. *American Economic Review*, 91(5), 1521–1538.

You, J. S. (2019). Teaching Mixed Strategy Equilibrium through a Classroom Experiment. *Business Education Innovation Journal*, 11(2), 65.