

# Gaming or Guessing: Mixing and best-responding in *Matching Pennies*

Timothy Dang

November 19, 2009

## JOB MARKET PAPER

### Abstract

A natural intuition for mixed strategies in competitive games is that players randomize to remain unpredictable, but this is a theoretically fragile notion. A player should only randomize between strategies if indifferent, and even then could choose a pure strategy or any disequilibrium mixed strategy. Various theories instead describe mixed strategies not due to random play, but rather to heterogeneous pure-strategy play. I conduct experiments in which game players are mirrored by guessers who make predictions about game play. This distinguishes best-responding (by the guessers) from game playing. In a *Matching Pennies* game, I find that game players are both more interested in unpredictability and actually more random.

## 1 Introduction

In traditional theories of simultaneous games, each player treats the others as part of the environment, essentially as probability distributions over states (strategies). A game player in these models isn't psychologically playing a game so much as engaging in an individual decision problem, with a possibly complex environment. In games with mixed strategy solutions, this traditional view of other players creates a bit of a paradox. Mixed strategies are a core element of noncooperative game theory, and the origins of the notion of mixed strategies are in individual randomization:

In playing Matching Pennies against an at least moderately intelligent opponent, the player will not attempt to find out the opponent's intentions but will concentrate on avoiding having his own intentions found out, by playing irregularly "heads" and "tails" in successive games. (von Neumann and Morgenstern, 1953, p.144)

However, it is hard to explain why an individual player should randomize. Harsanyi (1973) pointed out an inconsistency between randomization in mixed strategy Nash equilibrium (MSNE) and the traditional view of other

players as part of the environment, writing, “equilibrium points in mixed strategies are unstable because any player can deviate without penalty from his equilibrium strategy even if all other players stick to theirs.”

Harsanyi then “rescued” the solution concept by showing with his purification theorem that the same aggregate behavior can arise in a pure strategy Bayesian equilibrium of a slightly disturbed game. Other solution concepts such as quantal response equilibrium (McKelvey and Palfrey, 1995) and level- $k$  reasoning (Nagel, 1995; Stahl and Wilson, 1995) are typically described as heterogeneous behavior in pure strategies. In all these solution concepts, it would be difficult for an outside observer to distinguish active randomization from heterogeneous play<sup>1</sup>.

Some theories support individual randomization. The first type formalizes deliberate unpredictability. Under such theories, player A plays *as if* Player B might see and react to A’s strategy choice. Reny and Robson (2004) create a model in which players have heterogeneous beliefs about how likely it is that their strategy choices are observable. Explicitly re-adding the predictability element to the model causes some randomizing strategies (such as minimax in *Matching Pennies*) to be strictly preferred. The heterogeneous beliefs create a situation analogous to Harsanyi’s, in which other mixed strategies (such as the equilibrium strategies of a coordination game) are entirely due to heterogeneous pure strategies.

The second type of theory incorporates ambiguity aversion (also called *uncertainty aversion*). Ambiguity is uncertainty that can’t be captured by a subjective probability distribution. Ambiguity-averse decision makers are uncomfortable with unknowns, and can behave in ways contrary to expected utility theory. Just as an individual uncertain of how many red and black balls are in an urn might avoid betting on either (Ellsberg, 1961), a game player who is uncertain about their opponent’s strategy might not want to commit to a single responding strategy. By randomizing their move, a player can hedge against making a poor choice<sup>2</sup>.

Luce and Raiffa (1957) draw attention to these dual explanations of mixed strategies as ambiguity-averse behavior (which can make sense even in games against nature), or the secrecy gained by randomizing:

...it may be sensible to use a mixed strategy as a hedge against extremely unfavorable situations and against the possibility that one’s opponent has more insight into one’s behavior than anticipated. (Luce and Raiffa, 1957, p.73)

---

<sup>1</sup>In a different approach, Aumann and Brandenburger (1995) describe Nash equilibrium as an equilibrium in *beliefs* about others’ strategies, rather than in actual strategies. Since the beliefs need not be correct, actual play may be in pure strategies.

<sup>2</sup>Only some models of ambiguity aversion lead to deliberate randomization. Eichberger and Kelsey (1996) argue that more plausible models do not imply randomization.

In spite of the theoretical move away from individual mixing, experimental and field studies of mixed strategies are typically framed in terms of individual randomization and deliberate unpredictability. The question asked is whether players are good enough randomizers to match equilibrium probability distributions in frequency and with independence in repeated games.

Until recently, most results have not supported the ability of players to do so. Rapoport and Budescu (1992) found players in *Matching Pennies* switched between strategies too much. Brown and Rosenthal (1990) re-examined results from the 4x4 zero-sum game in O’Neill (1987) and rejected minimax play in both frequencies and serial independence.

Failure to play MSNE strategies could be due to a disinterest in playing such strategies, or poor ability to randomize properly in spite of the desire. Bar-Hillel and Wagenaar (1991) surveys studies of individual production of random sequences, and identify robust deviations from iid production. Subjects over-alternate (as in Rapoport and Budescu (1992)) and over-balance frequencies over short regions<sup>3</sup>. Shachat (2002) allowed explicit mixed strategies in O’Neill’s game to try to separate the two. While almost 75% of moves were explicitly mixed, most mixtures were not minimax, and there was still serial correlation of actions.

Noussair and Willinger (2003) allowed similar explicit mixing in a 2x2 “unprofitable game” with distinct MSNE and maximin solutions, as well as a cooperative solution with much greater payoffs. Many moves were randomized, although use declines from 60% of players in the first period to 41.5% in the final period. Game play was more consistent with QRE than MSNE, maximin, or cooperative play.

Ochs (1995) had subjects play *Matching Pennies* variants with many periods of random subject rematching—one “straight” zero-sum game and two *Asymmetric Matching Pennies* games. Ten games were played each period, and subjects defined at the beginning of the period how many of the games should be played with each strategy. The results were inconsistent with both MSNE and maximin play. McKelvey et al. (2000) also conducted experiments with *Asymmetric Matching Pennies*, without explicit mixing. Like Ochs, they found the overall frequencies of play inconsistent with MSNE.

The experiments with explicit randomization made it costless, so that a player who is indifferent among strategies might choose to mix. In Dang (2008), I report experiments in which indifference was made incompatible with a random strategy by charging a fee to explicitly randomize. In those experiments, few subjects chose to do so (approximately 10% of moves were explicitly mixed), and subjects’ self-described motivations primarily were consistent with purification. For those who were interested in randomization,

---

<sup>3</sup>Bar-Hillel and Wagenaar (1991) also look at judgement about random processes, which show similar biases.

there was not more interest in a repeated game than a true one-shot game. However, in the repeated game, mixtures did improve in the sense of increasing security level over the course of a 12-period experiment.

To examine ambiguity in games, Camerer and Karjalainen (1994) used a variant on *Matching Pennies* in which player 1 wished to mismatch, and player 2 wished to match. Player 1 had the standard **H** or **T** strategies. Player 2 had three strategies: **H**, **T** or randomize. In each of two treatments, more than half of player 2's chose to randomize. However, player 2 wasn't quite randomizing over pure strategies. Instead, the choice to randomize meant a 50-50 chance of each player winning *regardless* of Player 1's move<sup>4</sup>.

A number of field studies have been motivated by the idea that highly experienced players in high-stakes games may become expert in randomization. Walker and Wooders (2001) focus on the idea that, “[i]n many strategic situations it is important that one’s actions not be predictable by one’s opponent.” In a study of serves by professional tennis players, they found that win rates for left and right serves were consistent with minimax, implying the receiver was playing minimax. However, serves were negatively serially correlated rather than iid, as MSNE would predict. Chiappori et al. (2002) and Palacios-Huerta (2003) find that in the sub-game of a soccer penalty kick, professional soccer players both kick and block consistent with minimax, including serial independence of actions. Palacios-Huerta and Volij (2008) extend study of professionals by having professional soccer players play laboratory experiments, including the O’Neill game. In their analysis, it appears that expertise in the field translates to ability in the laboratory, with the soccer players playing according to minimax in both frequencies and independence. However, Wooders (2008) identifies non-stationary play and argues that student subjects actually played closer to minimax than the soccer players. Levitt et al. (2007) partially replicate the experiments with soccer players, in addition to world class poker and bridge players, and find that none play according to minimax.

Whether experts appear to be playing according to MSNE or not, studies of naturally-occurring data are necessarily weak in identifying the difference between randomization and heterogeneous play.

This paper reports two experiments that separate game-motivated mixing for unpredictability from individual-motivated mixing due to ambiguity aversion. The approach is motivated in part by Rubinstein’s criticism of randomization theories:

“[*Matching Pennies*] is classically used to motivate the notion of mixed strategy equilibrium, but randomization is a bizarre description of a player’s deliberate strategy in the game. A player’s action

---

<sup>4</sup>I am reformulating the game slightly from Camerer & Karjalainen’s presentation to make it easier to describe briefly.

is a response to his guess about the other player’s choice; guessing is a psychological operation that is very much deliberate and not random.” (Osborne and Rubinstein, 1994, p.37)<sup>5</sup>

In this experiment, pairs of players play repeated *Matching Pennies* games. Each player is mirrored by a “guesser” who receives similar payoffs for correctly guessing the other player’s actions. Differences in behavior between a player and their paired guesser reflect game-playing concerns with unpredictability. Players and guessers both have the option to explicitly randomize their actions using the computer. Since both types of subjects are in identical *guessing* circumstances, differences in use of the randomization device show how unpredictability motivates mixing by the players. Explicit randomization by guessers is motivated by ambiguity aversion, without the goal of unpredictability. While both types of subjects choose to explicitly randomize, game players do so significantly more than guessers. Players’ interest in being unpredictable carries over to actual unpredictability of actions. A probit test of predictability finds that players are less predictable than guessers.

The paper is organized as follows. Section 2 describes my experimental design. Section 3 describes the research questions in relation to the design. Experimental procedures are in section 4 and results are in section 5.

## 2 Experiment Design

I conduct an experiment using a standard repeated *Matching Pennies* game, with the participation of both players and guessers. This creates distinction between behavior that is game-oriented (such as trying to be unpredictable) and behavior that is essentially individual choice.

While the basic game is a two player game, it involves a quartet of subjects. Two subjects— $P^A$  and  $P^B$ , called “players”—play a competitive game. As usual in a two-player game, player  $P^A$ ’s payoff depends on their own action and the action of player  $P^B$ . Each of the players is paired with a “guesser”, who is given similar incentives. Player  $P^A$  is paired with guesser  $G^A$ , who is not truly a game player. Guesser  $G^A$  mirrors player  $P^A$  in that both are trying to best-respond to  $P^B$ ’s actions. Each period, guesser  $G^A$  is asked to predict what player  $P^B$  will do. The guesser receives the same payoff for a correct guess that the player receives for winning the game. So, guesser  $G^A$ ’s payoff depends on their own action (guess) and the action of player  $P^B$ . However,  $G^A$  is not part of the game—their action has no effect on the payoff of either player.

---

<sup>5</sup>This quote comes from Ariel Rubinstein, in a paragraph marked as disagreeing with his co-author. Rubinstein (1991) argues that the purification idea should only be used if the sources of the disturbances are identified in the game description.

Table 1: *Matching Pennies*

		$P^B$			
		<b>H</b>	<b>T</b>		
$P^A$	<b>H</b>	\$0.50	0	0	\$0.50
	<b>T</b>	0	\$0.50	\$0.50	0

Table 2: *Matching Pennies* payoff tables for guessers

		$P^B$		$G^B$	
		<b>H</b>	<b>T</b>	Guess <b>T</b>	Guess <b>H</b>
$G^A$	Guess <b>H</b>	\$0.50	-	0	-
	Guess <b>T</b>	0	-	\$0.50	-

$P^A$	<b>H</b>	-	0	-	\$0.50
	<b>T</b>	-	\$0.50	-	0

(a) Guesser  $G^A$

(b) Guesser  $G^B$

Table 1 shows the *Matching Pennies* game, with a low payoff of \$0.00 and a high payoff of \$0.50. Player  $P^A$  is the matcher—the row player—and player  $P^B$  is the mis-matcher—the column player. Table 2(a) shows the payoff table for guesser  $G^A$ , given their actions and those of player  $P^B$ .  $G^A$ 's payoffs depend on their action and  $P^B$ 's action in the same way as  $P^A$ 's payoffs in Table 1, but  $P^B$ 's payoffs are missing because the guesser's actions don't influence any player's payoffs.

Just as guesser  $G^A$  mirrors player  $P^A$ , guesser  $G^B$  mirrors player  $P^B$ . Table 2(b) shows the payoff table for guesser  $G^B$ , given their actions and those of player  $P^A$ . Again,  $G^B$ 's payoffs are the same as those for  $P^B$  in Table 1, and  $P^A$ 's payoffs are missing because the guesser's actions don't influence any player's payoffs. The difference in Table 2(b) is in the labeling of  $G^B$ 's strategies. Where the left column in Table 1 is labeled “**H**” for player  $P^B$ , the left column in Table 2(b) is labeled “Guess **T**” for guesser  $G^B$ . Thus, the guesser receives a payoff of \$0.50 for correctly guessing  $P^A$ 's action. Since player  $P^B$  is the mis-matcher,  $P^B$  playing **H** is equivalent to  $G^B$  guessing **T**, and vice-versa.

Information flows only one way. Guessers see the actions of both players, but, since guessers' actions don't affect players' payoffs, players aren't given any information about the moves of guessers<sup>6</sup>. The guessers also don't affect each others' payoffs and so are not shown each others' actions.

Since nothing they do can affect other subjects' payoffs or information,  $G^A$  and  $G^B$  are facing individual risky decision problems, although they are doing so with other players taking the role normally assigned to “Nature”. On the other hand,  $P^A$  and  $P^B$  are truly playing a game. While we don't know how  $P^A$  forms expectations about  $P^B$ 's strategy,  $G^A$  should form expectations

<sup>6</sup>Players were not informed that there were guessers in the experiment.

in the same way.

## 2.1 Repeated *Matching Pennies*

The subjects play 50 repeated games of *Matching Pennies*, as described above and shown in Table 1 with the same quartet. Players are able to play pure or explicitly randomized strategies, by having the computer “flip a coin”<sup>7</sup>. After each play, they see only the realized pure strategies before playing again.

The guessers were likewise able to choose either pure strategies (specific guesses) or explicitly mixed strategies (flipping the coin to make the guess).

## 2.2 Post-Experiment Questionnaire

A goal of this experiment is to get inside subjects’ heads, to identify motivations as well as behavior. To help answer these questions, at the end of the experiment there is a basic questionnaire of how and why they made the choices they did<sup>8</sup>.

The primary questions are open-ended. Subjects are asked, “Please take a moment and explain how you decided whether to flip the coin or not”. Responses are coded into a few (not exclusive) categories. Reasons to mix are categorized as “**Hedging**” (ambiguity aversion), or “**Counter prediction**” (deliberate unpredictability). Reasons to play an apparently pure strategy are categorized as “**Want to play pure**” for those who indicate a real preference for a pure strategy.

Finally, responses are coded if they indicate that the repeated-game aspect is important, as “**Repeated (confident)**” if they believe they can recognize the other player’s patterns or “**Repeated (unconfident)**” if they believe they can’t recognize the patterns, or they believe their opponent can recognize their own patterns better.

## 3 Research Questions

Unpredictability has a strong intuitive attraction as a motivation for mixing. However, the weight of theory goes against the desire to be unpredictable motivating random play. The essence of this experiment is to put a pair of players in the same situation, modulo the concern for predictability.

**Question 1:** Do game-players explicitly randomize more than guessers?

---

<sup>7</sup>For clarity, in the rest of this paper “flip” refers to using the computer as a randomization device, while “mix” or “randomize” can involve the computer or any mental or external method.

<sup>8</sup>In addition to the open-ended strategy questions, subjects are asked their gender and whether they have played similar games in the past. Responses to those questions are not addressed in this study.

The guesser  $G^A$  and player  $P^A$  both have the same information to form expectations, and the same money payoffs for correctly or incorrectly anticipating player  $P^B$ . From a guessing perspective, they are in the same situation. If concern with unpredictability is unimportant, then the player has no more motivation to randomize than the guesser. The null hypothesis that guessers and players explicitly randomize equally often will be tested against the one-sided alternative that the players will flip with greater frequency than the guessers, due to unpredictability motivations.

This leaves open the question about how much the guesser will randomize. Lack of experiments has left us uncertain about the implications of ambiguity on randomization<sup>9</sup>. Some models of ambiguity aversion imply a decision maker would randomize to hedge, while others do not. What is important here is that a player concerned with ambiguity and unpredictability has greater motivation to mix than one concerned with ambiguity alone.

Assuming that ambiguity does motivate randomization, a player with extreme ambiguity aversion will randomize in response to any expectations. A player with more moderate ambiguity aversion will play a pure strategy when confident of a strong likelihood that their opponent will play a particular strategy, but will randomize with lower confidence. Suppose in *Matching Pennies* that the matching player will play **H** whenever their subjective probability of their opponent playing **H** is  $p > \bar{p}$ . In that case their belief that their opponent will play **H** is great enough to overcome their uncertainty about the true probability. Similarly, the matcher will play **T** when  $p < \bar{p}$ . Then for  $p \in [\underline{p}, \bar{p}]$ , their uncertainty outweighs their beliefs in one or the other strategy, and the player will randomize as a hedge<sup>10</sup>.

Adding concern with unpredictability as in Reny and Robson (2004), suppose that the player also believes there is probability  $z$  that their opponent will be able to observe and respond to their strategy<sup>11</sup>. Now, the relevant thresholds become functions of the strategy chosen. When  $z > 0$ , the (subjective) probability that the opponent will play **H** when the matcher plays **H** is  $p(1 - z)$ —the probability of playing **H** *and* not having the opportunity to observe and respond. Similarly, the probability that the opponent will play **T** if the matcher plays **T** is  $(1 - p)(1 - z)$ .

So with concerns about predictability, the same player will randomize if  $p \in \left[ \frac{p-z}{1-z}, \frac{\bar{p}}{1-z} \right]$ . For  $z > 0$  and  $p \in (0, 1)$ , this interval is strictly larger

<sup>9</sup>The exception, as mentioned in Section 1, is (Camerer and Karjalainen, 1994) in which experiments with exactly this question in mind found some statistically insignificant evidence that increased ambiguity led to increased randomization.

<sup>10</sup>In a multiple-priors maximin expected utility model as in Gilboa and Schmeidler (1989), rather than a single subjective  $p$  being in the interval, the player will mix when any one of their priors falls in the interval. The essence of the argument remains the same.

<sup>11</sup>Unlike in more naturalistic settings such as war, sport, or business, in the laboratory it is clearly impossible for their opponent to observe their strategy. This probability of observation  $z$  can come from any concern a player has about their own predictability. In a repeated game, this could be the concern that the opponent will successfully identify a pattern of play.

than  $[p, \bar{p}]$ , so a player will be more likely to randomize with concerns about unpredictability than with ambiguity alone. If ambiguity aversion does not motivate randomization, then  $\underline{p} = \bar{p} = \frac{1}{2}$ , and unpredictability still enlarges the interval for which they will randomize.

**Question 2:** Are guessers and game-players equally predictable in their actions?

A player concerned with unpredictability would naturally be trying to be more unpredictable than a guesser unconcerned with unpredictability, so I want to find out if players are actually less predictable than guessers. However, experiments have shown people to be poor randomizers when they try to randomize. A player trying to be unpredictable by randomizing on their own may succumb to subconscious patterns and paradoxically become more predictable than the guesser. Therefore, this question will be addressed with two-sided tests.

“Predictability” can be a difficult to identify, since it requires assumptions about conditioning—predictable given *what*. As is common in studies of mixed strategies, I analyze predictability in terms of serial correlation with runs tests. A runs test can only identify predictability (non-iid play) conditioned on one’s own previous play. I also use a probit method described in (Noussair and Willinger, 2003), which looks for predictability conditioned on own and other-player moves and game outcomes.

## 4 Experiment Procedures

The experiments were conducted in November 2008 in the Economic Science Laboratory (ESL) at the University of Arizona. Subjects were undergraduates recruited from ESL’s subject database. The experiment had 55 subjects in three sessions<sup>12</sup>. Subjects were recruited for up to one and one-half hours, and the actual duration of the experiment from arrival to departure was approximately 1 hour.

The experimental games and questionnaires were conducted using z-Tree (Fischbacher, 2007), with computerized instructions using the same program. The instructions were strongly framed in terms of the player-guesser dichotomy. Players received all their instructions in terms of “play”, “lose”, “win”, and “opponent”. Guessers had their actions described in terms of “guess”, “wrong”, and “correct”<sup>13</sup>. In addition to their own instructions, guesser  $G^A$  saw all of player  $P^B$ ’s instructions (and  $G^B$  saw  $P^A$ ’s instructions),

<sup>12</sup>There was one group in which the  $G^B$  role was played as a dummy by the experimenter. Data generated by the experimenter and relevant comparisons are omitted from the analysis.

<sup>13</sup>Eliaz and Rubinstein (2008) conducted *Matching Pennies* experiments with framing of subjects’ roles. Subjects who moved second (but without any information advantage), and were asked to “guess” the first player’s move had an advantage.

so the guesser can understand the game from the player’s point of view. A summary of the rules remained on the screen during the experiment<sup>14</sup>.

Players were told, “You have two ways to decide whether to play Heads or Tails. (1) You can simply decide to play Heads or Tails. (2) You can flip the coin, then the computer will randomly play for you.” The instructions for flipping were similar for the guessers. During play, subjects had three buttons to choose each period whether to play (guess) **H** or **T**, or flip the coin.

Each of the subjects was randomly assigned to one of the subject types (player or guesser, and matcher or mis-matcher), and retained that type throughout the experiment. Subjects were in the same group for all 50 periods. Subjects had access to a history of play and outcomes.

At the end of each period, subjects were shown the play and outcome from that period. During play, subjects could also review a history which contained the same information:

**Players** were shown their own and their opponent’s realized moves, and the resulting payoff. They were not shown whether or not their opponent flipped the coin. In the history, a player could be reminded if they had flipped the coin themselves in a previous period. The players did not see any information about the guessers.

**Guessers** were shown their own guess, both players’ realized moves, and the resulting payoff. They were not shown whether or not the players flipped the coin, and they were not shown any information about the other guesser. In the history, a guesser could be reminded if they had flipped the coin themselves in a previous period.

At the end of the experiment, subjects were privately paid their earnings, plus a \$5.00 show-up fee. Average earnings (not including the show-up fee) were \$12.60.

## 5 Results

Table 3 gives descriptive statistics<sup>15</sup>. For the analysis, both player types will be pooled, as will both guesser types.

### 5.1 Explicit randomization

**Question 1:** Do game-players explicitly randomize more than guessers?

---

<sup>14</sup>Text of the instructions is in the appendix. The z-Tree treatments are available from the author.

<sup>15</sup>The presentation of the moves for  $P^B$  and  $G^B$  will vary depending on the context. Since three of the four subjects in a quartet— $P^A$ ,  $G^A$ , and  $G^B$ —wish to match, sometimes  $P^B$ ’s move will be inverted so that a move of **T** is treated as a guess of **H**, and vice-versa. Alternatively, sometimes  $G^B$ ’s guesses will be inverted so that a guess of **T** is presented as a move of **H**. This will be made explicit where either transformation is used.

Table 3: *Matching Pennies* Descriptive statistics

Variable	$P^A$	$G^A$	$P^B$	$G^B$	Players	Guessers
# <b>H</b>	24.71 (4.10)	28.21 (5.89)	23.71 (3.67)	25.38 (8.48)	24.21 (3.85)	26.85185 (7.25)
# Flips	14.14 (15.21)	9.79 (14.18)	21.71 (19.04)	10.69 (12.98)	17.93 (17.34)	10.22 (13.36)
Payoff	\$12.00 (\$0.90)	\$12.32 (\$1.81)	\$13.00 (\$0.90)	\$13.15 (\$1.38)	\$12.50 (\$1.02)	\$12.72 (\$1.64)
N	14	14	14	13	28	27

Standard errors are in parentheses.

“Players” is for  $P^A$  and  $P^B$  pooled, and “Guessers” is for  $G^A$  and  $G^B$  pooled.

$P^B$  moves have been transformed so a move of **T** is shown here as a guess of **H**.

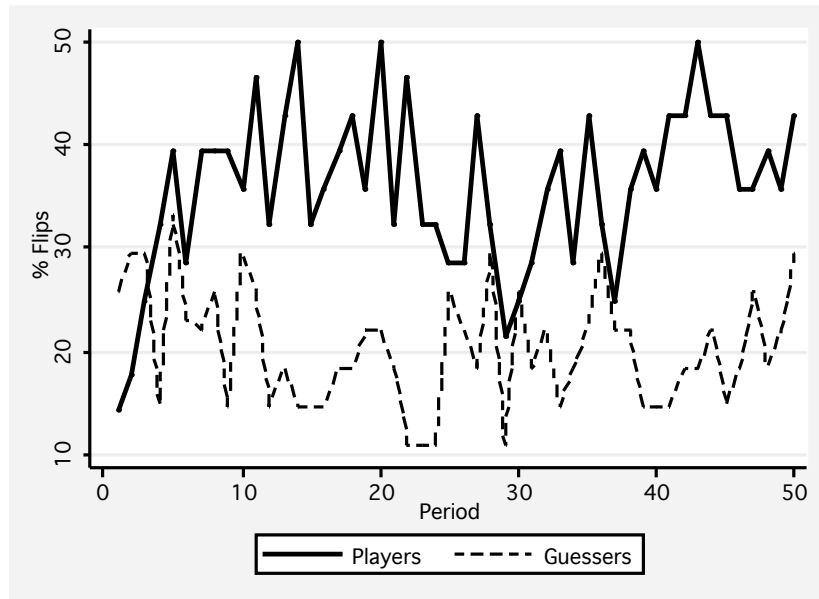


Figure 1: *Matching Pennies* Flipping per period

Players consistently choose to flip the coin to explicitly randomize more than guessers. Figure 1 shows the percentage of moves in each period in which a subject (player or guesser) decided to flip the coin to explicitly randomize. Players decided to flip more than guessers in 46 of the 50 periods. Overall, players flipped for approximately 35.6% of their moves, while guessers flipped for approximately 20.4% of theirs. At their maximum use, half of players flipped the coin in three periods, and in four periods as few as 11.1% of guessers flipped the coin. There is no trend for subjects to flip more or less over time, with the possible exception of increasing flipping by players in the first few periods.

Figure 2 shows a histogram of how many times each type of subject decided to flip. Many subjects of both types flipped few times, with 10 players and 12 guessers flipping five or fewer times<sup>16</sup>. However, the remaining guessers largely cluster with fewer than 20 flips and a median of 7 flips per guesser, while players were more evenly distributed with a median of 13.5 flips per player.

A bootstrap test of the mean difference between the number of times a player flipped versus their paired guesser rejects the null hypothesis of the same frequency at the 5% confidence level in favor of the alternative hypothesis that players flip more. There is no significant correlation between the frequency with which a player and their paired guesser ( $P^A$  and  $G^A$  or  $P^B$  and  $G^B$ ) decided to flip the coin<sup>17</sup>.

## 5.2 Predictability

Unpredictability in game-play could arise from many sources. Randomization due to ambiguity aversion would create unpredictability even if that isn't the goal. A player may be unpredictable while playing pure strategies if they have a random or unstable method for choosing which mixed strategy to play. These experiments are intended to distinguish deliberate unpredictability from these other sources.

Players demonstrate a concern with unpredictability by explicitly randomizing more than guessers. This interest in unpredictability carries over to actual unpredictability. One measure of unpredictability shows game players to be less predictable than guessers, although the traditional runs test doesn't identify a difference. The runs tests find that both types of subjects have positive serial correlation, and so are somewhat predictable.

With the explicit randomization device, I can sometimes observe for certain that a subject chose to be random. I cannot observe for certain that any sub-

<sup>16</sup>Four players and three guessers decided not to flip at all

<sup>17</sup>A Kendall's Tau test fails to reject independence with  $p=0.801$ .

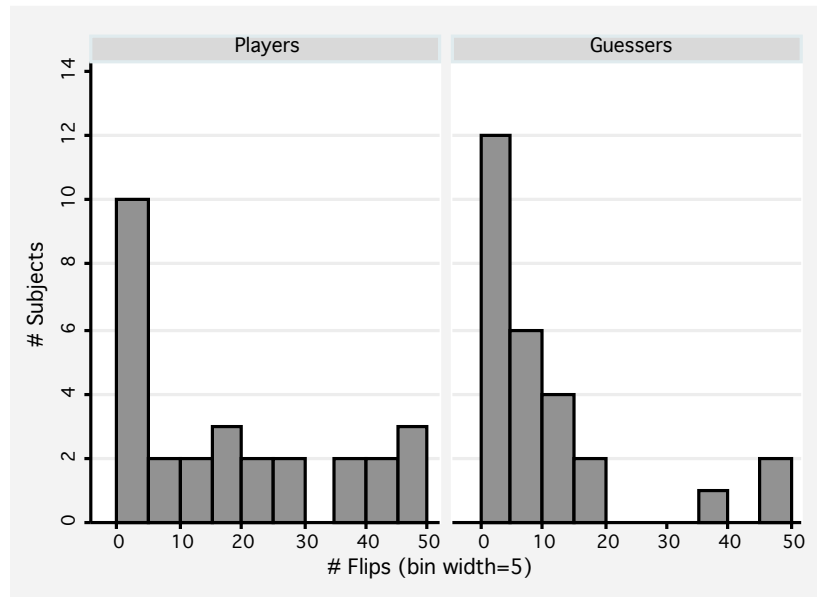


Figure 2: Histogram: *Matching Pennies* Number of flips per subject

ject chooses a pure strategy, since subjects may be randomizing mentally<sup>18</sup>. To study randomness, I will treat explicit and unobserved randomization equally by only considering realized actions.

**Question 2:** Are guessers and game-players equally predictable in their actions?

I take two approaches to measuring unpredictability. First, I use a runs test to look at whether any subjects’ mixed-strategy play is iid. A runs test looks at the number of “runs” in a series of data. For instance, a player who played **H**, **H**, **T**, **H**, **T**, **T** over six moves would have 4 runs ([**H**, **H**], [**T**], [**H**], and [**T**, **T**]). For iid play, there is an expected number of runs. Fewer runs indicates positive serial correlation—a player tends to repeat a particular pure strategy. More runs indicates negative serial correlation—a player tends to switch from one strategy to the other.

I want to see if the number of runs per subject has the predicted overall distribution. For instance, the expected number of runs is 26 if play is iid 50% **H**. If every subject had exactly 26 runs, the runs test would not reject iid play for any subject. Overall, however, it would be clear that something other than true iid play was occurring. For instance, if play was truly iid, approximately 16% of subjects should have fewer than 22 runs. I follow

<sup>18</sup>Alternatively, subjects might be using some unobservable external randomization device, such as poker-champion Dan Harrington’s watch (Harrington and Robertie, 2004, pp.52–53).

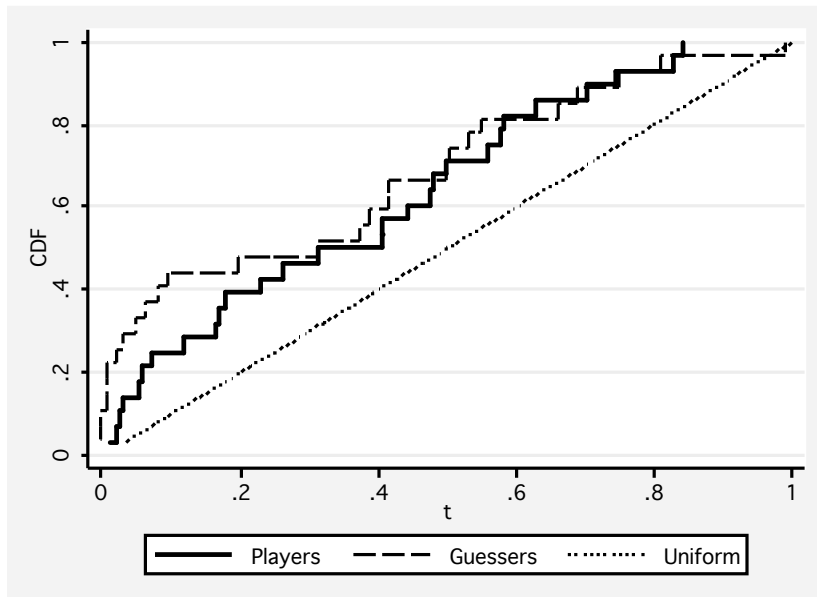


Figure 3: *Matching Pennies* t-statistics from runs tests

Walker and Wooders (2001), considering the entire distribution of results from the runs tests. From each runs test, I generate a t-statistic, a randomized statistic roughly equivalent to the lower-tailed p-value. Unlike the p-value, the t-statistic is continuously and uniformly distributed according to the null of iid play.

Figure 3 shows the CDF's of the t-statistic. According to the null, the distribution should be close to the uniform 45-degree line. Since the t-statistic comes from one-sided p-values, a distribution above the 45-degree line indicates too few runs, and so positive serial correlation of moves<sup>19</sup>. The further above the 45-degree line, the more subjects had a small p-value in the test for too few runs. According to the runs test, both kinds of subject showed a degree of predictability. Game players and guessers show positive serial correlation<sup>20</sup>. This positive correlation is counter to typical results for individual randomization tasks and some mixed strategy experiments, and is discussed in Section 6.

Kolmogorov-Smirnov (KS) tests on the distribution of the t-statistics reject uniform distribution at a 10% confidence level, but do not reject similar distribution for game players and guessers at a 10% confidence level<sup>21</sup>.

<sup>19</sup>Conversely, a distribution of the t-statistic below the 45-degree line would indicate negative serial correlation, but that did not occur in these experiments.

<sup>20</sup>The number of runs and t-statistics for each subject are shown in Tables 4-7

<sup>21</sup>Comparing the distribution to uniform, the KS test for players has  $p = 0.081$ , while guessers have  $p < 0.01$ . Comparing players to guessers,  $p > 0.30$ .

A player who merely echoes whatever their opponent did in the previous period (playing **H** in period  $t$  if and only if the opponent played **H** in period  $t - 1$ ) may have an iid distribution as identified by a runs test (depending on what their opponent does), but is also completely predictable. The second approach to identifying predictability is a technique used by Noussair and Willinger (2003) which can identify predictability conditioned on history of both other- and own-moves.

For each subject, I conducted the probit estimation given in Equation 1 of the probability of choosing **H** given the history of player moves<sup>22</sup>. The variable  $UL_{t-1}$  is a dummy indicating the outcome in the previous period was (Up, Left)—i.e. both players played **H**.  $UL3_{t-1}$  combines the previous three periods,  $UL3_{t-1} = UL_{t-1} + UL_{t-2} + UL_{t-3}$ . The  $DL$  and  $UR$  variables are similar for their respective outcomes, and  $DR$  (**T**, **T**) is omitted to avoid colinearity. The one-period lags allow for immediate response, while the three-period lags allow for response to trends<sup>23</sup>. The estimation results were ignored, except that they were used to find the pseudo- $R^2$  goodness-of-fit measure. This  $R^2$ , between 0 and 1, is used as a measure of the predictability of a player’s decisions.

$$p_t = \Phi(\alpha + \beta_1 UL_{t-1} + \beta_2 DL_{t-1} + \beta_3 UR_{t-1} + \gamma_1 UL3_{t-1} + \gamma_2 DL3_{t-1} + \gamma_3 UR3_{t-1}) \quad (1)$$

Figure 4 shows the empirical CDF’s of the  $R^2$ ’s derived from the probit regressions, for both players and guessers. Also shown is a baseline derived from simulating MSNE play<sup>24</sup>. A CDF below this baseline has generally higher pseudo- $R^2$  values, showing greater predictability. With this approach, the players’ CDF closely traces the baseline, with a maximum  $R^2$  of 0.288, while guessers are more predictable, with a maximum  $R^2$  of 0.635. A bootstrap test of the mean difference between the predictability (as measured by the pseudo- $R^2$ ) of a player versus their paired guesser rejects the null of equal predictability at the 5% confidence level<sup>25</sup>. Subjects who decided to flip the coin did achieve greater randomness. Figure 5 shows subjects’  $R^2$  versus the number of times they flipped. An OLS regression gives a rough correlation. Guessers are more scattered, with many higher  $R^2$  values, particularly for those who rarely flipped. On the right-hand side of the plot, subjects who flipped regularly show low predictability.

In the analysis of predictability in their experiment, Noussair and Willinger (2003) argue that unpredictability is attributable to error in play rather than

<sup>22</sup>Only player moves were used as independent variables in the regression, for both player and guesser moves as dependent variables. This allows easier comparison between player and guesser predictability, but gives a conservative measure of guesser predictability since their own history is absent.

<sup>23</sup>Noussair and Willinger (2003) used only the single-period lagged dummy variables.

<sup>24</sup>The baseline distribution was derived from 2000 iterations of play by a simulated quartet, with both independent and dependent variables being Nash play of iid 50% chance of playing (guessing) **H**.

<sup>25</sup>KS tests find similar results. Tests of subjects’ predictability versus the baseline reject similarity for the guessers ( $p = 0.000$ ), but not for the players ( $p = 0.470$ ). A KS test of the players’  $R^2$  versus the guessers’  $R^2$  rejects similarity with  $p = 0.027$ .

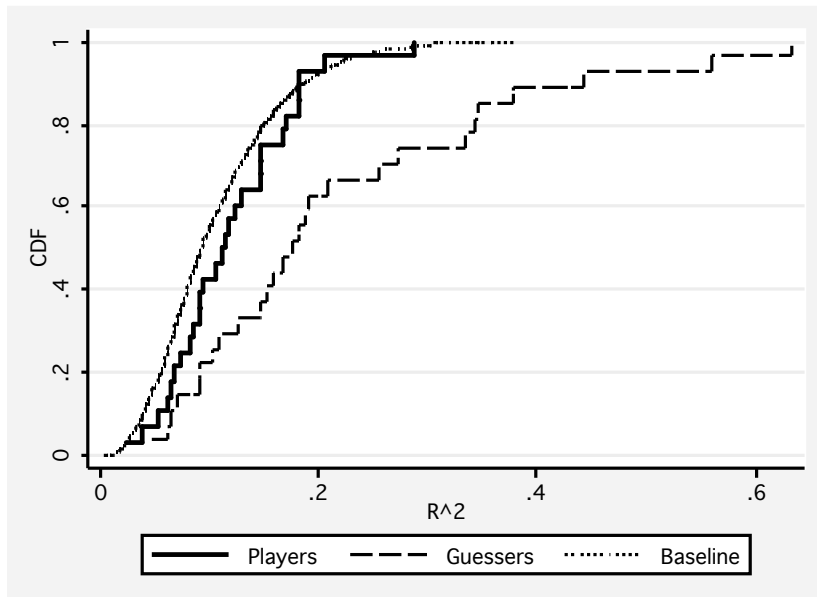


Figure 4: *Matching Pennies* Pseudo- $R^2$  from probit regressions

deliberate attempts to keep an opponent guessing. Consistent with QRE, they find that players are more predictable when they face a larger difference in expected payoffs between strategies (with the expectation coming from historical play). According to the predominant behavioral notion of QRE, players evaluate their payoffs from various pure strategies with (unbiased) error, and then play their subjectively best pure strategy. If the true payoffs from two strategies are close, the error is more likely to shift preferences to the inferior strategy than if one payoff is much larger.

In my experiments, own-payoff considerations cannot explain differences in predictability, since players and guessers faced the same own-payoff environment. The difference in predictability between players and guessers is—like the difference in interest in flipping the coin—attributable to the desire to be unpredictable.

### 5.3 Questionnaire responses

When the experiment was complete, subjects were asked to briefly describe their decisions while their payments were prepared. Subjects were asked how they decided whether or not to flip the coin. The responses were coded, looking for clear indications of whether or not a subject wanted to play randomly, and if so, why<sup>26</sup>.

<sup>26</sup>Complete questionnaire responses are in the appendix.

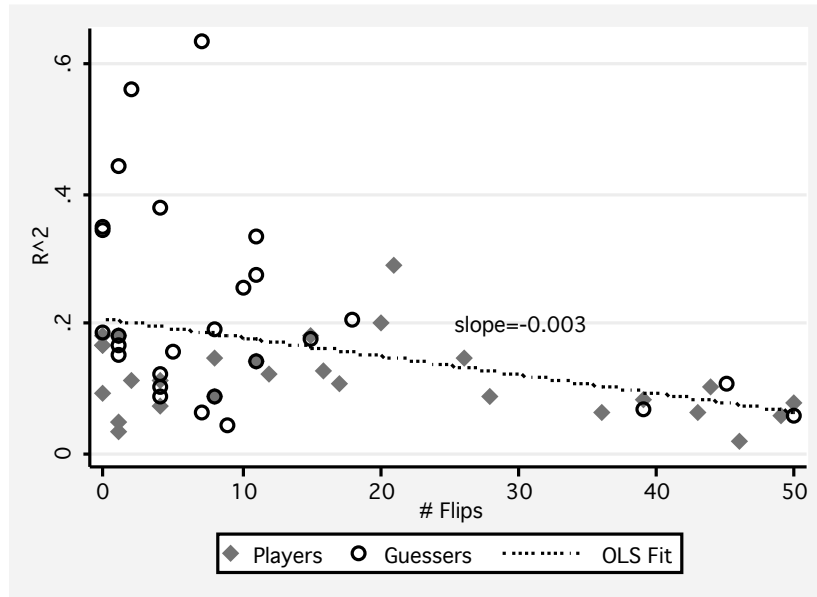


Figure 5: *Matching Pennies* Number of flips vs. probit pseudo- $R^2$

Table 4: *Matching Pennies* Matching Players

Subject	Flips	H	T	Binomial p	Runs	Runs t	Probit $R^2$
1	28	28	22	0.420	22	0.117	0.091
5	1	28	22	0.358	26	0.582	0.180
9	4	28	22	0.350	22	0.167	0.116
16	15	18	32	0.050	21	0.177	0.182
20	0	24	26	0.773	20	0.056	0.181
21	26	21	29	0.232	23	0.230	0.147
25	39	23	27	0.585	29	0.829	0.085
29	46	30	20	0.160	25	0.498	0.022
34	2	29	21	0.281	18	0.013	0.115
38	8	25	25	0.901	19	0.026	0.147
44	1	17	33	0.020	23	0.476	0.052
48	0	22	28	0.380	24	0.314	0.170
49	16	28	22	0.391	21	0.073	0.127
53	12	25	25	0.964	24	0.264	0.123

Table 5: *Matching Pennies* Mis-matching Players

Subject	Flips	H	T	Binomial p	Runs	Runs t	Probit $R^2$
2	44	25	25	0.984	20	0.057	0.105
6	1	23	27	0.555	27	0.578	0.037
10	49	20	30	0.137	27	0.704	0.062
12	36	24	26	0.859	26	0.479	0.065
17	17	20	30	0.181	28	0.843	0.110
22	8	32	18	0.045	25	0.627	0.090
26	43	20	30	0.181	26	0.559	0.063
30	50	24	26	0.798	25	0.404	0.082
32	0	27	23	0.507	19	0.020	0.166
35	4	26	24	0.885	25	0.407	0.072
45	20	20	30	0.121	25	0.445	0.204
50	11	26	24	0.736	20	0.033	0.144
54	0	26	24	0.865	23	0.171	0.093

Table 6: *Matching Pennies* Matching Guessers

Subject	Flips	H	T	Binomial p	Runs	Runs t	Probit $R^2$
3	0	23	27	0.602	28	0.748	0.342
7	4	28	22	0.468	20	0.061	0.090
11	7	27	23	0.535	25	0.374	0.064
14	4	34	16	0.011	12	0.000	0.379
18	4	30	20	0.187	28	0.810	0.124
23	0	26	24	0.717	6	0.000	0.188
27	2	45	5	0.000	2	0.000	0.560
31	11	28	22	0.335	26	0.502	0.145
36	11	27	23	0.520	21	0.081	0.335
40	45	25	25	0.937	26	0.532	0.109
42	1	23	27	0.670	25	0.386	0.167
46	8	23	27	0.651	26	0.547	0.090
51	1	24	26	0.755	21	0.094	0.443
55	39	32	18	0.064	31	0.991	0.069

Table 7: *Matching Pennies* Mis-matching Guessers

Subject	Flips	H	T	Binomial p	Runs	Runs t	Probit $R^2$
4	1	20	30	0.157	27	0.686	0.180
8	5	25	25	0.904	26	0.498	0.158
13	9	21	29	0.253	18	0.022	0.046
15	7	44	6	0.000	11	0.416	0.635
19	50	27	23	0.608	27	0.659	0.061
24	4	33	17	0.021	18	0.048	0.101
28	8	14	36	0.002	20	0.313	0.191
33	15	31	19	0.072	22	0.197	0.175
37	10	26	24	0.725	25	0.415	0.255
43	11	21	29	0.298	17	0.009	0.274
47	0	11	39	0.000	12	0.008	0.347
52	1	29	21	0.225	17	0.006	0.150
56	18	28	22	0.369	19	0.030	0.207

Table 8: *Matching Pennies* Coded responses

	$P^A$	$G^A$	$P^B$	$G^B$
Want to play pure	22.6%	22.6%	11.3%	15.1%
Want to be random	17.0%	9.4%	17.0%	15.1%
Hedging	3.8%	5.7%	9.4%	9.4%
Counter prediction	0.0%	0.0%	1.9%	1.9%
Repeated (confident)	9.4%	7.5%	5.7%	13.2%
Repeated (unconfident)	0.0%	0.0%	1.9%	0.0%
Lazy	5.7%	3.8%	1.9%	1.9%

Table 8 shows the percentage of subjects indicating each motivation in their decisions<sup>27</sup>. Responses were coded based on whether the subject indicated an interest in being random, and whether their randomness was motivated by the desire to be unpredictable or to hedge against bad outcomes. If they mentioned the repeated-game aspect, that was coded based on whether the subject was unconfident (they were afraid they would show recognizable patterns) or confident (they were trying to exploit their opponent’s recognizable patterns). The “**Lazy**” category was added after reviewing the questionnaire responses. If a subject indicated more than one motivation (random sometimes, pure other times, for instance), all were included.

There is no clear indication of whether subjects wanted to play mixed or pure strategies.  $P^A$  and  $G^A$  subjects were more likely to indicate a desire to play a pure strategy, but  $P^B$  and  $G^B$  subjects were more likely to indicate a desire to be random. Responses are more consistent in regards to *why* a subject might want to play a pure or mixed strategy. Players expressed more interest in being random than did guessers, and  $P^A$  and  $P^B$  subjects were both more likely to indicate that randomization was motivated by ambiguity aversion than by countering the opponent’s prediction<sup>28</sup>. Likewise, the players were more likely to express confidence that they could recognize their opponents’ patterns in the repeated game than fear of their own patterns being exploited.

## 6 Discussion and Conclusions

This paper reports an experiment meant to separate simple best-responding from game-playing in a competitive game. Contrary intuitions about what is natural in games has led to different stories supporting mixed-strategy solutions. Under the classic interpretation, a mixed strategy is a player randomizing their action. However, since such randomization is difficult to rationalize—whether players are indifferent over the pure strategies or not—alternative notions of mixed-strategy play have different players playing dif-

<sup>27</sup>The percentages shown are based on non-blank responses, and categories are not exclusive.

<sup>28</sup>One  $G^B$  subject clearly misunderstood the experiment, indicating a desire to be unpredictable.

ferent pure strategies. The situation is further complicated by the possibility that people randomize due to ambiguity aversion, whether in a game or facing an individual decision.

By mirroring each game player with a guesser who has the same incentives to best-respond, but no interest in being unpredictable, I can distinguish randomization with the intent to be unpredictable from individual best-response behavior. In a repeated *Matching Pennies* game, players were both more interested in random play, and actually more random. Since players and guessers were in mostly similar circumstances, the additional randomness by players can be attributed to their desire to be unpredictable.

The most common method to identify randomness in individual actions or game play has been the runs test. Here, both subject types have positive serial correlation as identified by too few runs (too many long runs). Bar-Hillel and Wagenaar (1991) describe too few long runs and too much alternation as, “fairly robust findings ... that have since withstood the test of time,” for psychology experiments in generating random sequences. This tendency to negative serial correlation has been seen in mixed-strategy experiments as well, including O’Neill (1987) and Rapoport and Budescu (1992), as well as tennis field data (Walker and Wooders, 2001).

There have been other experiments with positive correlation. Rosenthal et al. (2003) find positive serial correlation for many subjects in “hide and seek” games, as does Okano (2008) in O’Neill’s game, and Slonim et al. (2003) in a variety of 2x2 games. The last speculate that face-to-face play, as in O’Neill (1987) may predispose players to alternating for unpredictability, as opposed to computerized experiments. That speculation has not been directly addressed by any experiment, but this experiment did use an anonymous computer interface and so is consistent with that explanation.

While runs tests identified all subject types as deviating from MSNE play, for these experiments the tests are unable to identify the difference in predictability between players and guessers. A probit analysis that includes more history—including the opponent’s moves and game outcomes—is able to show that guessers are more predictable than players.

Responses to the open-ended questionnaire give some idea of how subjects were thinking about their decisions. In the questionnaire following the experiment, players didn’t put a great deal of value on unpredictability. Both players and guessers were more likely to express an interest in playing a pure strategy than randomization. Both also expressed a greater concern with randomization for hedging than for unpredictability, and were more likely to believe they could exploit their opponents’ regularities in repeated play than to believe the opponents could exploit their own. Intuitively, this fits with the pattern of player and guesser coin-flipping. Game players chose to

flip the coin more than guessers, but less than twice as often. It appears that ambiguity aversion motivated a good deal—perhaps the majority—of randomization.

A challenging inference from the results of these experiments is that a holistic theory of mixed-strategy play may be behaviorally implausible. There is evidence of explicit randomization to be unpredictable, but there is also evidence of randomization to reduce risk, and heterogeneous pure-strategy play.

## References

- Aumann, R. and A. Brandenburger (1995). Epistemic conditions for nash equilibrium. *Econometrica* 63(5), 1161–1180.
- Bar-Hillel, M. and W. A. Wagenaar (1991). The perception of randomness. *Advances in Applied Mathematics* 12(4), 428–454.
- Brown, J. and R. Rosenthal (1990). Testing the minimax hypothesis: A re-examination of o’neill’s game experiment. *Econometrica* 58(5), 1065–1081.
- Camerer, C. F. and R. Karjalainen (1994). Ambiguity-aversion and non-additive beliefs in non-cooperative games: Experimental evidence. In B. Munier and M. J. Machina (Eds.), *Models and Experiments in Risk and Rationality*. Boston: Kluwer Academic Publishers.
- Chiappori, P., S. Levitt, and T. Groseclose (2002). Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *The American Economic Review* 92(4), 1138–1151.
- Dang, T. (2008, May). The meaning of mixing: Individual strategic randomization in games. Working paper.
- Eichberger, J. and D. Kelsey (1996). Uncertainty aversion and preference for randomisation. *Journal of Economic Theory* 7(1), 31–43.
- Eliasz, K. and A. Rubinstein (2008, July). Edgar allan poe’s riddle: Framing effects in repeated matching pennies games. Working paper.
- Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics* 75(4), 643–669.
- Fischbacher, U. (2007). z-tree - zurich toolbox for readymade economic experiments. *Experimental Economics* 10(2), 171–178.
- Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2), 141–153.

- Harrington, D. and B. Robertie (2004). *Harrington on Hold'em*. Henderson, NV: Two Plus Two.
- Harsanyi, J. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory* 2(1), 1–23.
- Levitt, S. D., J. A. List, and D. H. Reiley (2007, August). What happens in the field stays in the field: Professionals do not play minimax in laboratory experiments. Working paper.
- Luce, R. D. and H. Raiffa (1957). *Games and decisions. Introduction and critical survey*. New York: John Wiley & Sons, Inc.
- McKelvey, R. D. and T. Palfrey (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior* 10, 6–38.
- McKelvey, R. D., T. Palfrey, and R. Weber (2000). The Effects of Payoff Magnitude and Heterogeneity on Behavior in 2x2 Games with Unique Mixed Strategy Equilibria. *Journal of Economic Behavior and Organization* 42(4), 523–548.
- Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review* 85(5), 1313–1326.
- Noussair, C. and M. Willinger (2003, November). Efficient mixing and unpredictability in an experimental game. Working paper.
- Ochs, J. (1995). Games with unique mixed strategy equilibria: An experimental study. *Games and Economic Behavior* 10(1), 202–217.
- Okano, Y. (2008, June). Minimax play by teams. Working paper.
- O’Neill, B. (1987). Nonmetric test of the minimax theory of two-person zerosum games. *Proceedings of the National Academy of Sciences* 84(7), 2106–2109.
- Osborne, M. J. and A. Rubinstein (1994). *A Course in Game Theory*. Cambridge, Massachusetts: MIT Press.
- Palacios-Huerta, I. (2003). Professionals play minimax. *Review of Economic Studies* 70(2), 395–415.
- Palacios-Huerta, I. and O. Volij (2008). Experientia docet: Professionals play minimax in laboratory experiments. *Econometrica* 76(1), 71–115.
- Rapoport, A. and D. Budescu (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General* 121(3), 352–363.

- Reny, P. J. and A. J. Robson (2004). Reinterpreting mixed strategy equilibria: a unification of the classical and bayesian views. *Games and Economic Behavior* 48(2), 355–384.
- Rosenthal, R. W., J. Shachat, and M. Walker (2003). Hide and seek in arizona. *International Journal of Game Theory* 32(2), 273–293.
- Rubinstein, A. (1991). Comments on the interpretation of game theory. *Econometrica* 59(4), 909–924.
- Shachat, J. (2002). Mixed strategy play and the minimax hypothesis. *Journal of Economic Theory* 104(1), 189–226.
- Slonim, R. L., I. Erev, and A. E. Roth (2003, July). Regularities in play of zero sum games with unique mixed strategy equilibria. Working paper.
- Stahl, D. O. and P. W. Wilson (1995). On players’ models of other players: Theory and experimental evidence. *Games and Economic Behavior* 10(1), 218–254.
- von Neumann, J. and O. Morgenstern (1953). *Theory of Games and Economic Behavior* (3rd ed.). Princeton University Press.
- Walker, M. and J. Wooders (2001). Minimax play at wimbledon. *The American Economic Review* 91(5), 1521–1538.
- Wooders, J. (2008, June). Does experience teach? professionals and does experience teach? professionals and minimax play in the lab. University of Arizona Department of Economics Working Paper Series 2008.