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

Timothy Dang

May 15, 2009

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. In an *Asymmetric Matching Pennies* game, I look at whether players are willing to forgo expected payoff in order to be unpredictable, and find little difference between players and guessers, with players being somewhat better at exploiting disequilibrium play.

1 Introduction

In traditional theories of simultaneous games, each player treats the others as part of the environment, in this case 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[1].

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[2].

Luce and Raiffa (1957) draw attention to this dual explanations of mixed strategies as amiguity-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

[1]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.

[2]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 over-switching in *Matching Pennies*. 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[3]. 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 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,

[3]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[4].

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, or by one's opponents." 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-occuring 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

[4]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)[5]

In these experiments, pairs of players play 2x2 games (either *Matching Pennies* or *Asymmetric Matching Pennies*). 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.

In the first experiment, pairs of game players play a repeated *Matching Pennies* game. 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.

If mixing to be unpredictable is valuable to game players, they should be willing to trade-off some other objective with unpredictability. To see if players will trade-off expected payoff with unpredictability, the second experiment has subjects play a series of *Asymmetric Matching Pennies* games, rematched with a new opponent every period. Subjects only have pure strategy actions available. Previous experiments with this game have had consistent results away from MSNE, so there have been exploitation opportunities which can best be exploited with pure strategies. Differences in behavior between player and guesser would be consistent with a disequilibrium willingness to forgo an exploitation opportunity in order to be unpredictable. In fact, there is little difference between player and guesser behavior, with players being slightly better than guessers at exploiting disequilibrium play. Both kinds of subjects still effectively play a mixture, although both the probit test and runs tests find players and guessers to be far from iid play.

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

[5] 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.

2 Experiment Design

I conduct two experiments, one using a standard repeated *Matching Pennies* game, and one using *Asymmetric Matching Pennies* with random rematching. The common element to both experiments is 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^Amirrors player P^Ain that both are trying to best-respond to P^B's actions. Each period, guesser $G^A$is asked to predict what player P^Bwill 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.

Table 1: *Matching Pennies*

		P^B			
		H		**T**	
P^A	**H**	\$0.50	0	0	\$0.50
	T	0	\$0.50	\$0.50	0

Table 2: *Matching Pennies* payoff tables for guessers

		P^B			
		H		**T**	
G^A	Guess **H**	\$0.50	-	0	-
	Guess **T**	0	-	\$0.50	-

(a) Guesser G^A

		G^B			
		Guess **T**		Guess **H**	
P^A	**H**	-	0	-	\$0.50
	T	-	\$0.50	-	0

(b) Guesser G^B

In both experiments, I use competitive games to avoid other-regarding preferences creating differences between players and guessers[6]. Table 1 shows the *Matching Pennies* game used in the first experiment, 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 3(a) shows the

[6]This attempt to avoid other-regarding preferences was imperfect. According to their questionnaire response, one mis-matching player in the *Asymmetric Matching Pennies* experiment deliberately played **H** frequently to give the matchers a shot at the high payoff.

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^Amirrors player P^A, guesser G^Bmirrors player P^B. Table 3(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^Bin 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 3(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 3(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^Bplaying **H** is equivalent to G^Bguessing **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[7]. 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^Aforms expectations about P^B's strategy, G^Ashould form expectations in the same way.

2.1 Repeated *Matching Pennies*

In the first experiment, the game is traditional constant-sum *Matching Pennies*, as described above and shown in Table 1. The subjects play 50 repeated games with the same quartet. Players are able to play pure or explicitly randomized strategies, by having the computer "flip a coin"[8]. 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). Unlike in Dang (2008).

[7]Players were not informed that there were guessers in the experiment.

[8]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.

2.2 *Asymmetric Matching Pennies*

The goal of the second experiment is to see the influence that randomization behavior has on gross game behavior. A game player seeking to randomize for unpredictability may choose to play a more random strategy than if they were solely concerned with best-responding[9]. This could appear not only in period-to-period independence of actions, but also in actions being played with different overall frequencies.

Straight *Matching Pennies* isn't well-suited to distinguish such gross differences in game behavior, since the $\left(\frac{1}{2}\mathbf{H}, \frac{1}{2}\mathbf{T}\right)$ equilibrium is so behaviorally strong. There's unlikely to be a tradeoff decision between best-responding and unpredictability. Instead I use an *Asymmetric Matching Pennies* game which does have such a tradeoff. The game matrix for players is shown in Table 3, and the guesser's payoffs are shown in Table 4. This is a non-constant-sum game, with a unique MSNE in which the row player mixes $\left(\frac{1}{2}\mathbf{H}, \frac{1}{2}\mathbf{T}\right)$, and the column player mixes $\left(\frac{1}{5}\mathbf{H}, \frac{4}{5}\mathbf{T}\right)$.

Table 3: *Asymmetric Matching Pennies*

	H		**T**	
H	$1.20	0	0	$0.30
T	0	$0.30	$0.30	0

Table 4: *Asymmetric Matching Pennies* payoff tables for guessers

		P^B			
		H		**T**	
G^A	Guess **H**	$1.20	-	0	-
	Guess **T**	0	-	$0.30	-

(a) Guesser G^A

		G^B			
		Guess **T**		Guess **H**	
P^A	**H**	-	0	-	$0.30
	T	-	$0.30	-	0

(b) Guesser G^B

There are two other differences between this experiment and the *Matching Pennies* experiment[10]. First, the same properties which make *Asymmetric Matching Pennies* good for identifying tradeoffs between unpredictability and best-responding raise the possibility that repeated-game strategies will be different than those for the stage-game[11]. Since the guessers aren't true game-players, and can't influence game-player behavior, such repeated-game concerns are irrelevant to them. To keep comparability between players and

[9]For my current purposes, exactly how to decide whether one strategy is "more random" than another can be left undefined. What is important is that a different strategy may be preferred because it is *subjectively* more random. I will assume that any non-degenerately mixed strategy is more random than a pure strategy.

[10]These two experiments should not be considered *treatments*. Firstly, there are too many differences between the experiments for controlled comparison. Also, each experiment has its treatments built-in—the players and guessers are two treatments running synchronously.

[11]Wooders and Shachat (2002) show that, for a zero-sum game with binary payoffs such as straight *Matching Pennies*, the repeated-game equilibrium is the same as the stage-game equilibrium.

guessers, in this experiment, players were randomly re-matched each period. While each player is re-matched to play against a new player every period, each player-guesser pair remains constant, so that in a given (P^A, G^A) pair, the player and guesser see the same history of game behavior. The random re-matching corresponds to previous experiments on *Asymmetric Matching Pennies* such as Ochs (1995) and McKelvey et al. (2000).

The final difference between the two experiments is that in the *Asymmetric Matching Pennies* experiment, there is no facility for explicit randomization. Each player and guesser is given only the strategies play (or guess) **H** or **T**. Since with this experiment, I want to see realized gross difference in game play, I want any randomization to be natural and "home grown".

2.3 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[12].

The primary questions are open-ended. For the *Matching Pennies* experiment, 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.

For the *Asymmetric Matching Pennies* experiment, subjects are asked, "Please take a moment and explain how you decided whether to play (guess) Heads or Tails." Responses are coded into similar categories, including "**Hedging**", "**Counter prediction**", and "**Want to play pure**". Since subjects are randomly re-matched, responses are not coded in regards to the repeated game. Instead of "**Random**", responses are coded for "**Non-stationary**" when a subject indicated they wanted variability in their strategy. Because of the random re-matching of subjects, a non-stationary strategy is effectively random. For instance, a player could choose **H**, **H**, **T**, **H**, **H**, **T**, ..., but for the other subejcts this would be equivalent to a $\frac{2}{3}$ **H** mixture. "**Non-stationary**" in

[12]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.

this context does *not* include regime-switching such as deciding in the second half of the experiment to use a different strategy than in the first half.

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?

The guesser $G^A$and player P^Aboth 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[13]. 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 $p < \underline{p}$. Then for $p \in \left[\underline{p}, \bar{p}\right]$, their uncertainty outweighs their beliefs in one or the other strategy, and the player will randomize as a hedge[14].

Adding concern with unpredictability as in Reny and Robson (2004), suppose that the player also believes there is probability z that their opponent

[13]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.

[14]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.

will be able to observe and respond to their strategy[15]. 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 $p \in \left[\frac{\underline{p}-z}{1-z}, \frac{\bar{p}}{1-z}\right]$. For $z > 0$ and $p \in (0, 1)$, this interval is strictly larger than $\left[\underline{p}, \bar{p}\right]$, 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.

Question 3: Are guessers better-responders than players?

If game players value randomization, they should be willing to make some tradeoffs between randomness and other objectives. In a situation where the best response is a predictable pure strategy, a desire for randomization could induce non-best-response mixed strategies.

In many games (such as *Asymmetric Matching Pennies* shown in Table 3), experiments have shown players to deviate systematically from mixed-strategy Nash equilibrium. In *Asymmetric Matching Pennies*, Ochs (1995)

[15]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 is could be the concern that the opponent will successfully identify a pattern of play.

and McKelvey et al. (2000) found that both row and column players overplay **H** relative to MSNE[16]. The deviation from equilibrium play makes pure strategies best responses. The row players' best response is pure-strategy **H**, while the column players' best-response is pure-strategy **T**. However, experimental subjects continue to place non-trivial weight on their inferior strategies.

To examine whether game players are willing to sacrifice payoff for unpredictability, quartets of subjects played *Asymmetric Matching Pennies*. In this experiment, the subjects were not provided a randomization device. Subjects were randomly re-matched each period, but a player-matcher pair was constant (i.e. a given player $P^B$and guesser G^Bwere in the same quartet each period), so the player and guesser will have seen the same history of moves in previous periods.

In a given period, the guesser $G^A$and player P^Awill both be in the same position in terms of best-responding to the expected strategy of player P^B. If concern with unpredictability is unimportant, then the player has no more motivation to randomize than the guesser. Since it is likely that players deviate systematically from MSNE, players and guessers could both benefit from exploiting that deviation.

The null hypothesis that guessers and players best-respond with the same frequency will be tested against the one-sided alternative that guessers will play the "exploiting" strategy with greater frequency than the players.

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 *Matching Pennies* experiment had 55 subjects in three sessions, and the *Asymmetric Matching Pennies* experiment had 47 subjects in two sessions[17]. 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

[16]Both Ochs (1995) and McKelvey et al. (2000) consider various *Matching Pennies* variations. The version of *Asymmetric Matching Pennies* in my experiment is similar to *game 3* in (Ochs, 1995) and *game D* in (McKelvey et al., 2000).

[17]In each experiment, there was one group in which the G^Brole was played as a dummy by the experimenter. Data generated by the experimenter and relevant comparisons are omitted from the analysis.

"guess", "wrong", and "correct"[18]. In addition to their own instructions, guesser G^Asaw all of player P^B's instructions (and G^Bsaw P^A's instructions), 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[19].

In the *Matching Pennies* experiment, 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. In the *Asymmetric Matching Pennies* experiment, there was no option to 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.

In the repeated *Matching Pennies* experiment, subjects were in the same group for all 50 periods. In the *Asymmetric Matching Pennies* experiment, game players were randomly re-matched each period and never played the same opponent two periods in a row. Guessers were paired constantly with a specific player for all 50 periods. In both experiments, 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:

***Matching Pennies* 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.

***Matching Pennies* 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.

***Asymmetric Matching Pennies* players** were shown their own and their opponent's moves, and the resulting payoff. The players did not see any information about the guessers.

***Asymmetric Matching Pennies* guessers** were shown their own guess,

[18]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.

[19]Text of the instructions is in the appendix. The z-Tree treatments are available from the author.

Table 5: ***Matching Pennies*** Descriptive statistics

Variable	P^A	G^A	P^B	G^B	Players	Guessers
# **H**	24.71	28.21	23.71	25.38	24.21	26.85185
	(4.10)	(5.89)	(3.67)	(8.48)	(3.85)	(7.25)
# Flips	14.14	9.79	21.71	10.69	17.93	10.22
	(15.21)	(14.18)	(19.04)	(12.98)	(17.34)	(13.36)
Payoff	$12.00	$12.32	$13.00	$13.15	$12.50	$12.72
	($0.90)	($1.81)	($0.90)	($1.38)	($1.02)	($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**.

Table 6: ***Asymmetric Matching Pennies*** Descriptive statistics

Variable	P^A	G^A	P^B	G^B
# **H**	28.08	24.00	35.75	33.18
	(10.77)	(11.35)	(8.68)	(5.46)
Payoff	$14.43	$13.65	$7.85	$7.72
	($3.93)	($3.08)	($1.08)	($0.85)
N	12	12	12	11

Standard errors are in parentheses.
P^B moves have been transformed so a move of **T** is shown here as a guess of **H**.

both players' moves, and the resulting payoff. They were not shown any information about the other guesser.

At the end of the experiment, subjects were privately paid their earnings, plus a $5.00 show-up fee. In the *Matching Pennies* experiment, average earnings (not including the show-up fee) were $12.60. In the *Asymmetric Matching Pennies* experiment, average earnings varied by role from $7.72 for G^B to $14.43 for P^A.

5 Results

Tables 5 and 6 give descriptive statistics for the *Matching Pennies* and *Asymmetric Matching Pennies* experiments respectively[20]. For the analysis of the *Matching Pennies* experiment, both player types will be pooled, as will both guesser types. Because of the asymmetry, most of the analysis for the *Asymmetric Matching Pennies* experiment will treat the types separately.

[20]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.

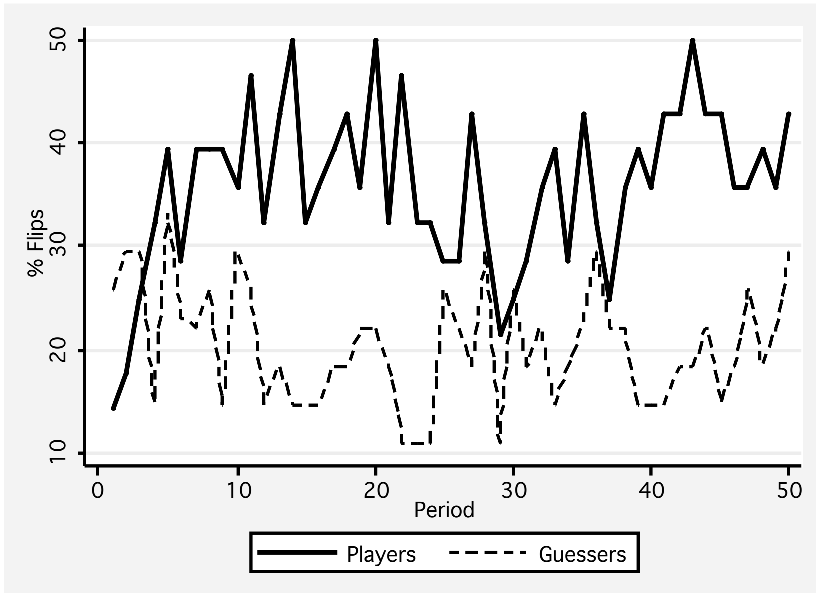

Figure 1: *Matching Pennies* Flipping per period

5.1 Explicit randomization

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

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[21]. 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.

[21] Four players and three guessers decided not to flip at all

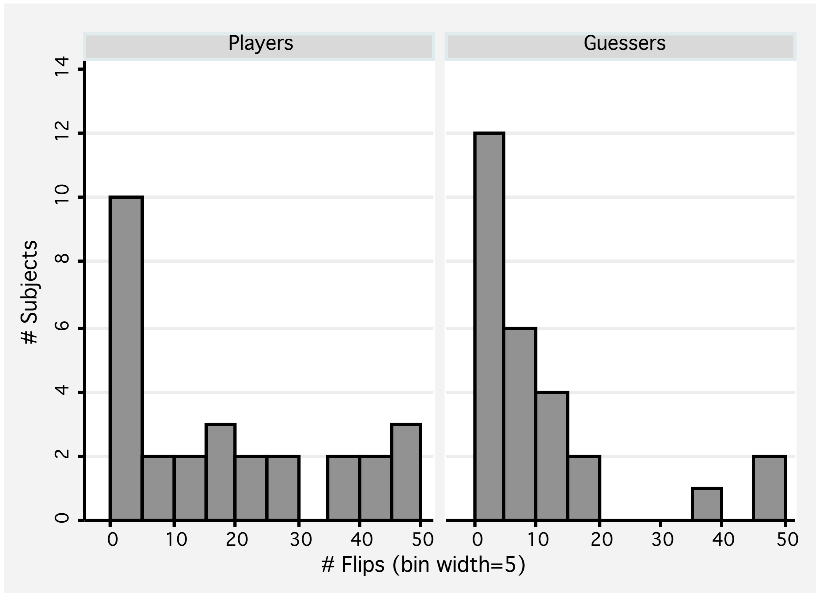

Figure 2: Histogram: *Matching Pennies* Number of flips per subject

---

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^Aor $P^B$and G^B) decided to flip the coin[22].

5.2 Predictablity

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.

In the *Matching Pennies* experiment, 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. In the *Asymmetric Matching Pennies* experiment with random re-matching of players, nei-

[22]A Kendall's Tau test fails to reject independence with p=0.801.

ther measure of predictability finds a difference between players and guessers. The runs tests find that both types of subjects in both experiments have positive serial correlation, and so are somewhat predictable.

With the explicit randomization device in the *Matching Pennies* experiment, I can sometimes observe for certain that a subject chose to be random. I cannot observe for certain that any subject chooses a pure strategy, since subjects may be randomizing mentally[23]. 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 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 for the *Matching Pennies* experiment, and Figures 4 and 5 show CDF's of the t-statistic for *Asymmetric Matching Pennies* matchers and mis-matchers, respectively. 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[24]. The further above the 45-degree line, the more subjects had a small p-value in

[23] 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).

[24] 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.

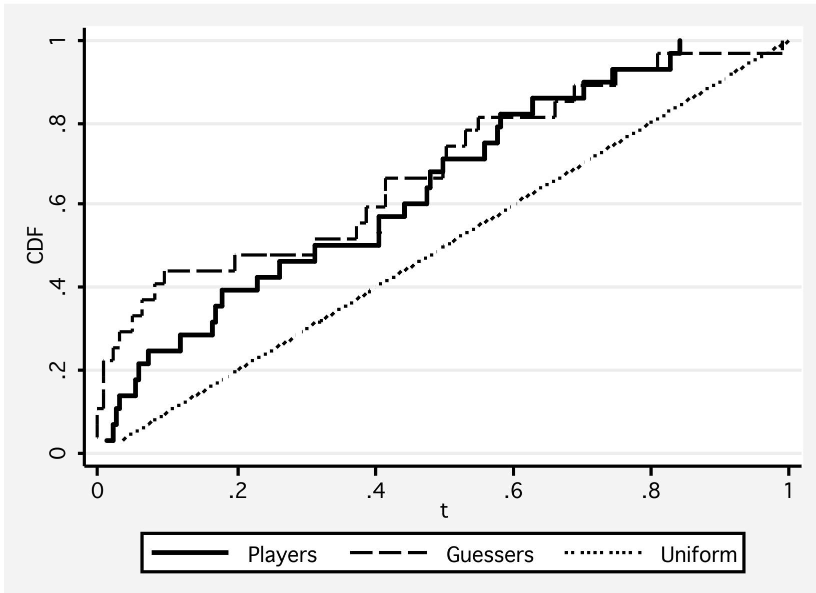

Figure 3: *Matching Pennies* t-statistics from runs tests

---

the test for too few runs. According to the runs test, every kind of subject shows a degree of predictability. In both experiments, game players and guessers show positive serial correlation[25]. 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 in both experiments (and for both matchers and mis-matchers separately in *Asymmetric Matching Pennies*) reject uniform distribution at a 10% confidence level, but do not reject similar distribution for game players and guessers at a 10% confidence level[26][27].

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

[25]The number of runs and t-statistics for each subject are shown in Tables 7-10 for the *Asymmetric Matching Pennies* experiment, and Tables 11-14 for the *Asymmetric Matching Pennies* experiment.

[26]Comparing the distribution to uniform, the KS test for *Matching Pennies* players has $p = 0.081$. *Matching Pennies* guessers and all *Asymmetric Matching Pennies* subjects have $p < 0.01$. Comparing players to guessers, each group has $p > 0.30$.

[27]For the *Asymmetric Matching Pennies* experiment, since there were two sessions with randomly re-matched subjects, many subjects interacted during the experiment. Because of the nature of the null, statistical tests against hypotheses of MSNE play are reported without qualification. For tests comparing players to guessers, reported p-values should be considered lower bounds on possible p-values in the case of correlation between subjects.

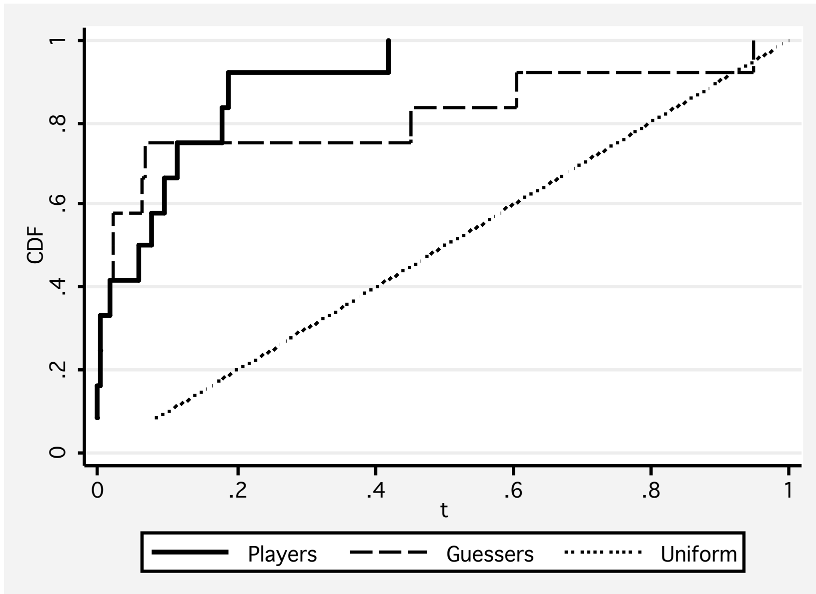

Figure 4: *Asymmetric Matching Pennies* P^A and G^A t-statistics from runs tests

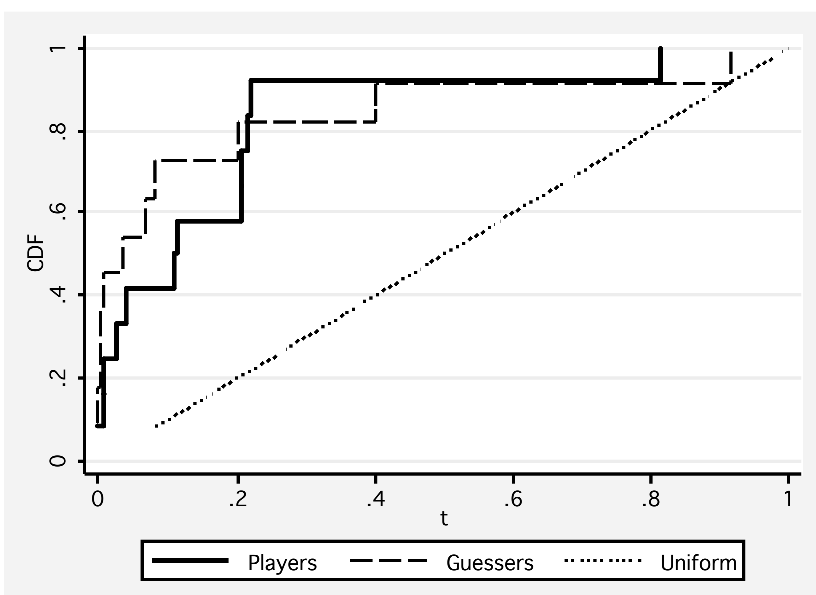

Figure 5: *Asymmetric Matching Pennies* P^B and G^B t-statistics from runs tests

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[28]. 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[29]. 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.

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

For the *Matching Pennies* experiment, Figure 6 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[30]. 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 pseduo-R^2) of a player versus their paired guesser rejects the null of equal predictability at the 5% confidence level[31]. Subjects who decided to flip the coin did achieve greater randomness. Figure 9 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 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

[28]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.

[29]Noussair and Willinger (2003) used only the single-period lagged dummy variables.

[30]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**.

[31]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 $p = 0.027$.

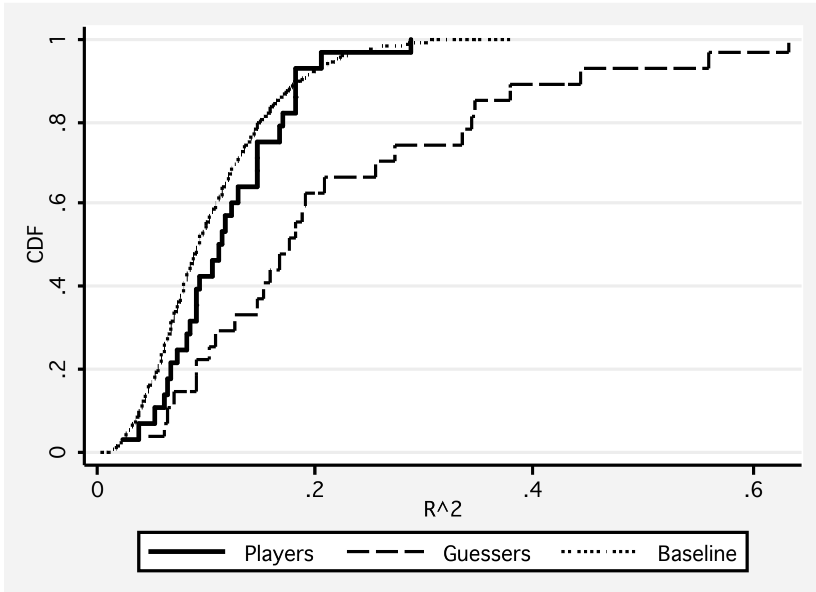

Figure 6: *Matching Pennies* Pseudo-R^2 from probit regressions

---

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.

In the *Asymmetric Matching Pennies* experiment, there is no clear difference in predictability between players and guessers. Figures 7 and 8 show the empirical CDF's of the R^2's, for matchers and mis-matchers respectively, as well as a MSNE baseline[32]. In Figure 8, the mis-matching guessers (G^B) CDF is below the baseline, indicating greater predictability, but otherwise appears to follow the shape of the baseline. However, it is not statistically distinct from the mis-matching players (P^B). The G^Aguessers and both player types are more distant from the baseline. A bootstrap test of the mean difference in predictability (with matchers and mis-matchers considered separately) fails to reject equal predictability at the 10% confidence level[33].

[32] The baseline distributions for *Asymmetric Matching Pennies* were derived similarly to those for *Matching Pennies*, except that the probabilities of **H** for mis-matching players and guessers were the MSNE 20% chance of playing **H** (guessing **T**).

[33] KS tests find similar results. Tests of subjects' predictability versus the baseline reject similarity for all types at the 1% confidence level. KS tests of the players' R^2 versus the guessers R^2 fails to reject similarity (p = 0.854 for matchers, and $p = 0.397$ for mis-matchers.).

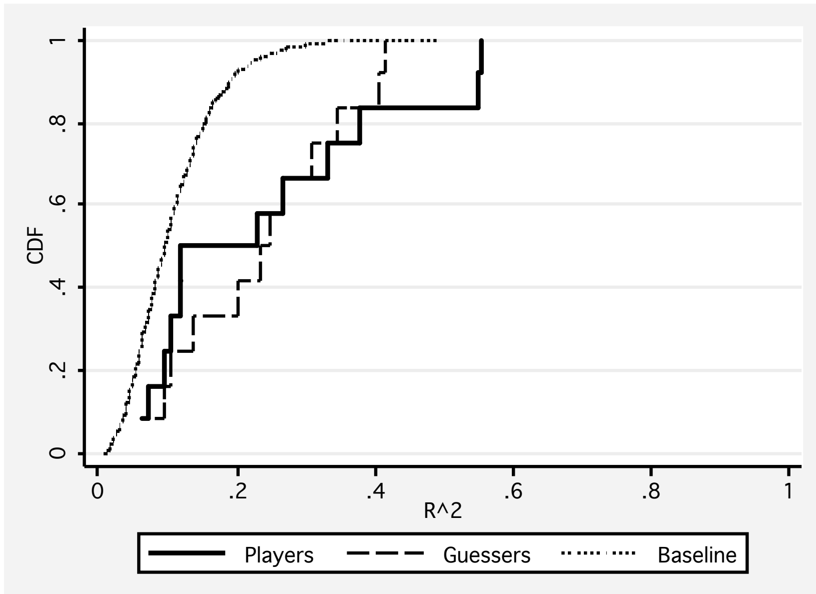

Figure 7: *Asymmetric Matching Pennies* Matchers' pseudo-R^2 from probit regressions

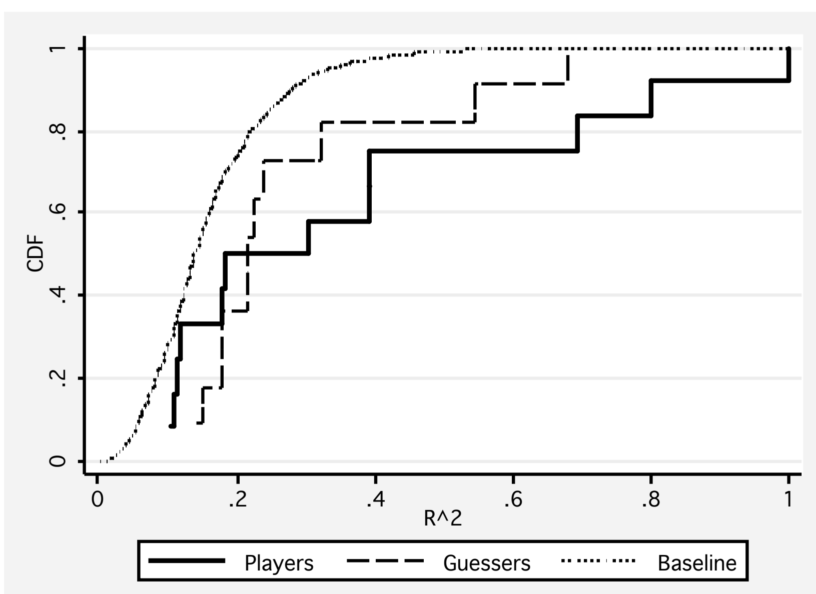

Figure 8: *Asymmetric Matching Pennies* Mis-matchers' pseudo-R^2 from probit regressions

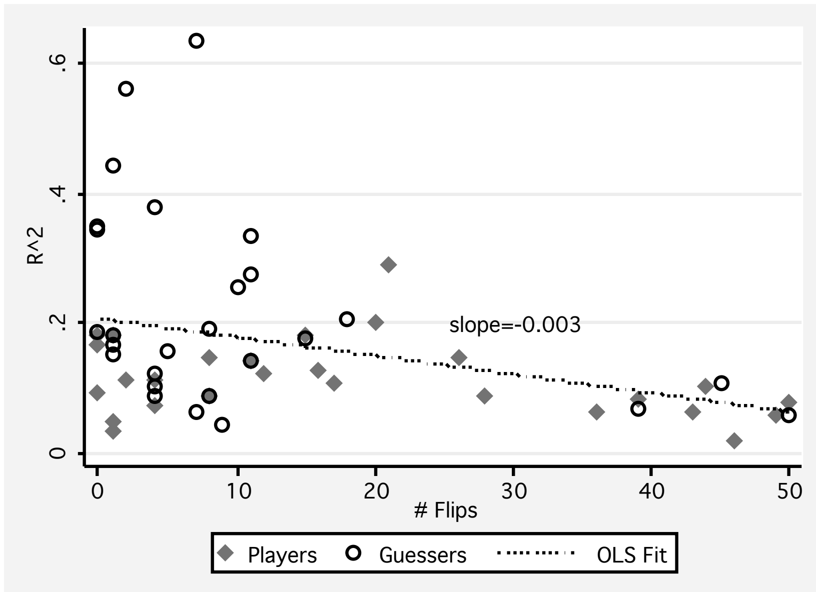

Figure 9: *Matching Pennies* Number of flips vs. probit pseudo-R^2

Table 7: *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 8: *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 9: *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 10: *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 11: *Asymmetric Matching Pennies* Matching Players

Subject	**H**	**T**	**Binomial p**	**Runs**	**Runs t**	**Probit** R^2
57	26	24	0.734	21	0.060	0.550
61	35	15	0.004	21	0.421	0.105
65	26	24	0.870	21	0.097	0.231
72	36	14	0.002	8	0.000	0.553
76	16	34	0.015	20	0.189	0.119
77	50	0	0.000	1	0.000	0.119
81	43	7	0.000	9	0.015	0.379
85	15	35	0.004	19	0.115	0.098
90	21	29	0.279	15	0.001	0.269
94	23	27	0.589	15	0.001	0.333
100	21	29	0.314	22	0.179	0.071
104	25	25	0.973	21	0.077	0.065

Table 12: *Asymmetric Matching Pennies* Mis-matching Players

Subject	**H**	**T**	**Binomial p**	**Runs**	**Runs t**	**Probit** R^2
58	40	10	0.843	13	0.028	0.304
62	46	4	0.018	3	0.000	1.000
66	18	32	0.000	21	0.208	0.182
68	31	19	0.005	22	0.214	0.105
73	31	19	0.002	16	0.007	0.392
78	28	22	0.000	29	0.812	0.120
82	46	4	0.016	6	0.038	0.798
86	47	3	0.011	5	0.113	0.691
88	41	9	0.819	10	0.006	0.391
91	31	19	0.005	20	0.109	0.108
101	36	14	0.137	19	0.219	0.180

Table 13: *Asymmetric Matching Pennies* Matching Guessers

Subject	**H**	**T**	**Binomial p**	**Runs**	**Runs t**	**Probit** R^2
59	11	39	0.000	11	0.003	0.200
63	32	18	0.035	12	0.000	0.309
67	26	24	0.696	19	0.021	0.266
70	15	35	0.004	23	0.605	0.416
74	20	30	0.201	25	0.450	0.406
79	13	37	0.000	16	0.068	0.095
83	23	27	0.609	21	0.064	0.084
87	19	31	0.105	14	0.001	0.107
92	13	37	0.000	15	0.023	0.135
96	28	22	0.431	31	0.947	0.249
98	44	6	0.000	6	0.001	0.234
102	44	6	0.000	8	0.019	0.344

Table 14: *Asymmetric Matching Pennies* Mis-matching Guessers

Subject	H	T	Binomial p	Runs	Runs t	Probit R^2
60	33	17	0.018	14	0.002	0.226
64	34	16	0.051	27	0.917	0.178
69	43	7	0.315	11	0.083	0.323
71	27	23	0.000	23	0.200	0.153
75	27	23	0.000	17	0.006	0.142
80	29	21	0.000	15	0.001	0.237
84	35	15	0.071	17	0.036	0.546
89	35	15	0.090	21	0.400	0.177
93	31	19	0.004	13	0.000	0.217
99	42	8	0.461	11	0.069	0.677
103	29	21	0.000	13	0.000	0.214

5.3 Frequency of pure strategies

Question 3: Are guessers better-responders than players?

Table 15 shows the frequency of decisions in the *Asymmetric Matching Pennies* game. As in previous experiments, both row and column players over-played **H** relative to MSNE. This created an opportunity for exploitation by playing a pure strategy. Given the row players' deviations, column players could earn an expected \$0.17 by playing **T**, versus \$0.13 from playing **H**. Row players could exploit column players' deviations by playing pure **H** for an expected \$0.34 versus \$0.21 for **T**.

If game players were willing to sacrifice some expected payoff for the unpredictability that comes from randomness, this could explain some playing of the lower-payoff strategies with positive probability. Guessers would receive no benefit from being unpredictable, however. Therefore, if players choose their strategies partly for unpredictability, they should be less responsive to the exploitation opportunities than guessers. In *Asymmetric Matching Pennies*, this would mean that G^Ashould play (guess) **H** with greater frequency than P^A, and that G^Bshould play **T** (actually guess **H**) with greater frequency than P^Bplays **T**.

Table 15 shows that, in overall frequencies, guessers are slightly *worse* than players at exploiting deviations from equilibrium. Approximately 56% of P^A's moves were the optimal **H**, versus G^A's 48%, and approximately 72% of P^B's moves were the optimal **T**, versus G^B's 67%. For 15 pairs, the player played the high-payoff pure strategy more often than the guesser, while only 6 pairs had the guesser playing the high-payoff strategy more.

For each subject, I performed a "randomized" binomial test on their frequency of playing (guessing) **H** versus the MSNE frequency. Similar to the t-statistic described in Section 5.2, the randomized binomial test gives a p-value which is continuously and uniformly distributed under the null hypoth-

Table 15: *Asymmetric Matching Pennies*

	H		**T**		EV	P^A	G^A
h	\$1.20	0	0	\$0.30	\$0.34	0.562	0.480
t	0	\$0.30	\$0.30	0	\$0.21	0.438	0.520
EV	\$0.13		\$0.17				
P^B	0.285		0.715				
G^B	0.334		0.666				

G^B guesses are transformed so a guess of **H** is shown as a move of **T**.

esis[34]. *Matching Pennies* players played close to the MSNE 50-50 frequency, with approximately 7% of rejecting MSNE play at the 5% significance level. Guessers were further from the MSNE frequency, with approximately 22% rejecting at the 5% confidence level, although a KS test on the distribution of p-values does not reject MSNE play for either players or guessers, and a two-sample KS test does not reject similarity of players and guessers.

In the *Asymmetric Matching Pennies* experiment, at least half of subjects—of each type—have MSNE play rejected by the binomial test at the 5% confidence level, and KS tests don't find differences in the distribution of p-values for players and guessers.

5.4 Questionnaire responses

When the experiment was complete, subjects were asked to briefly describe their decisions while their payments were prepared. For the *Matching Pennies* experiment, they were asked how they decided whether or not to flip the coin. For the *Asymmetric Matching Pennies* experiment, the subjects were asked how they decided to play **H** or **T**. The responses were coded, looking for clear indications of whether or not a subject wanted to play randomly, and if so, why[35].

5.4.1 Matching Pennies

Table 16 shows the percentage of subjects indicating each motivation in their decisions[36]. 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

[34] Unlike the t-statistic, this p-value is equivalent to a two-sided p-value. See Wooders (2008)

[35] Complete questionnaire responses are in the appendix.

[36] The percentages shown are based on non-blank responses, and categories are not exclusive.

Table 16: ***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%

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^Asubjects were more likely to indicate a desire to play a pure strategy, but $P^B$and G^Bsubjects 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^Bsubjects were both more likely to indicate that randomization was motivated by ambiguity aversion than by countering the opponent's prediction[37]. 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.

5.4.2 Asymmetric Matching Pennies

Table 17 shows the percentage of subjects indicating each motivation in their decisions. Rather than identifying random motivation, responses were coded for the desire to be **"Non-stationary"**. A strategy would be considered non-stationary if it deliberately varied the pure strategy action taken. This would include mixed strategies, but would not include a strategy which varied due to responding to variation by the opponent, or a subject changing their play during the course of the experiment.

Very few subjects indicated a desire to be random/non-strationary explicitly for reasons of hedging or unpredictability, so a meaningful comparison of which is more important is impossible. Every type of subject was more likely to indicate an interest in playing a pure strategy than a non-stationary one, and there was no clear difference between players and guessers in this regard.

[37]One G^Bsubject clearly misunderstood the experiment, indicating a desire to be unpredictable.

Table 17: ***Asymmetric Matching Pennies*** coded responses

	P^A	G^A	P^B	G^B
Want to play pure	15.9%	22.7%	25.0%	18.2%
Non-stationary	11.4%	11.4%	6.8%	11.4%
Hedging	0.0%	0.0%	0.0%	2.3%
Counter prediction	2.3%	0.0%	0.0%	0.0%

6 Discussion and Conclusions

This paper reports two experiments meant to separate simple best-responding from game-playing in competitive games. 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 different 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. In a randomly re-matched *Asymmetric Matching Pennies* game, players and guessers played similarly. Both varied their play over the available pure strategies, and neither successfully exploited potentially profitable deviation from MSNE play.

The most common method to identify randomness in individual actions or game play has been the runs test. Here, every subject type in both experiments has 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 *Matching Pennies* 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.

For the *Asymmetric Matching Pennies* experiment, subjects expressed little interest in being random (or even non-stationary) for any reason. Accordingly, players did not appear to trade-off expected payoff for unpredictability. In fact, players were somewhat better at exploiting disequilibrium play than guessers.

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.

Eliaz, 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.

Wooders, J. and J. M. Shachat (2002). On the irrelevance of risk attitudes in repeated two-outcome games. *Games and Economic Behavior 34*(2), 342–363.