The effect of matching mechanism on learning in games played under limited information

Nick Feltovich*
Department of Economics
Monash University
Clayton, VIC 3800, Australia
nicholas.feltovich@monash.edu

Sobei H. Oda
Faculty of Economics
Kyoto Sangyo University
oda@cc.kyoto-su.ac.jp

Abstract: We report results from a human-subjects experiment that examines the effect on learning of the mechanism used to match individuals to opponents in two-player games. Several games are played repeatedly under either a *fixed-pairs* or a *random matching* treatment. Unlike most economics experiments, the games are played under *limited information*: subjects are never shown a payoff matrix nor given information about opponent payoffs. We find that behavior in the treatments, while similar in early rounds, diverges over time. Depending on the game, the matching mechanism can affect which of multiple equilibria is likely to be reached, the speed of convergence, or other aspects of outcomes. Usually, fixed-pairs matching is associated with more likely coordination on a pure-strategy Nash equilibrium, more likely play of a higher-payoff Nash equilibrium, and faster convergence toward pure-strategy play. The differences we find are consistent with those implied by learning model simulations, and a follow-up experiment with a new game confirms our main findings.

* Corresponding author. Oda is supported in part by Grants-in-Aid of Scientific Research (#17310029 and #17700155) and by Japan's Ministry of Education, Culture, Sport, Science and Technology. Feltovich gratefully acknowledges support from the University of Aberdeen Business School. We thank Atsushi Iwasaki for conducting the experimental sessions, and Ido Erev for helpful comments.

1. Introduction

Colin Camerer (2003), summarizing the state of experimental game theory research in his behavioral game theory text, said, "[t]here are no interesting games in which subjects reach a predicted equilibrium immediately. And there are no games so complicated that subjects do not converge in the direction of equilibrium (perhaps quite close to it) with enough experience" (p. 20). The implication of this passage is that, while game-theoretic concepts like Nash equilibrium can be useful for characterizing how individuals behave after acquiring sufficient experience, a true behavioral game theory must incorporate a description of how individuals learn.[1] Though the beginnings of a learning theory are well established (Camerer & Ho, 1999; Fudenberg & Levine, 1998; Roth & Erev, 1995), we are far from a consensus regarding the exact nature of learning in games. In order for progress toward some kind of consensus to continue, more detailed study must be made of factors that can influence how decision makers learn.

The objective of this paper is to take a small step in this direction. We concentrate on the effects of one particular experimental design manipulation—the protocol used to match players to opponents—on the manner in which behavior changes over time. Many economics experiments comprise more rounds than there are potential opponents. For such experiments, there are two commonly-used matching mechanisms: *fixed pairs*, with players matched repeatedly to the same opponent, and *random matching*, with players randomly rematched after every round. We design and run an experiment in which human subjects repeatedly play each of six two-player games under one of these two matching mechanisms. In order to focus on learning, rather than other phenomena that might be sensitive to the matching protocol, as well as to minimize effects arising from other-regarding preferences, we give subjects only limited information about the games—in contrast to nearly all previous studies of matching mechanisms, which have focused on behavior under complete information (see Section 2.1). In our design, subjects are told they are playing a game, but are not given any information about payoffs before playing, and while they can learn about their own payoffs via end-of-round feedback, they never receive information about opponent payoffs.

Our results suggest that the matching mechanism can indeed have a sizable effect on observed outcomes in general, and learning in particular. Levels of cooperation are frequently higher under fixed pairs than under random matching. Also, convergence to an equilibrium is usually faster under fixed pairs, though in one game, the opposite is true. The matching mechanism also can have an effect on the likelihood of behavior converging to one equilibrium versus another: in one game, fixed-pairs matching actually increases the chance of players getting "stuck" on a

[1] Throughout this paper, we use the term "learning" to encompass any systematic change in behavior over time. This includes changes in play due to improved understanding of the structure of the game, as well as behavioral responses to changes in beliefs about how others play.

low-payoff equilibrium, while in the others, fixed pairs makes the high-payoff equilibrium more likely. The differences we find between fixed pairs and random matching are similar to those predicted by a modified version of Erev and Barron's (2005) RELACS learning model, the original of which has been shown previously to do a good job of characterizing behavior in repeated individual decision-making tasks under limited information. A follow-up experiment, in which we preserve the main features of our design but use a new game and change some of the experimental procedures, largely replicates the results of the original experiment, also consistently with the modified RELACS model.

2. The experiment

Figure 1 shows the games used in the original experiment. (The follow-up experiment is described in Section 5.)

Figure 1: Games used in the original experiment

		Player 2	
		C	D
Player 1	C	2,2	0,0
	D	0,0	1,1

Coordination Game (CG)

		Player 2	
		C	D
Player 1	C	0,0	3,5
	D	5,3	0,0

Battle of the Sexes (BoS)

		Player 2	
		C	D
Player 1	C	7,7	1,8
	D	8,1	4,4

Prisoners' Dilemma (PD)

		Player 2	
		C	D
Player 1	C	7,7	1,5
	D	5,1	5,5

Stag Hunt—high (SHH)

		Player 2	
		C	D
Player 1	C	5,5	–1,3
	D	3,–1	3,3

Stag Hunt—medium (SHM)

		Player 2	
		C	D
Player 1	C	1,1	–5,–1
	D	–1,–5	–1,–1

Stag Hunt—low (SHL)

While each game is symmetric and 2x2, they differ in some important ways. Prisoners' Dilemma has a strictly dominant strategy, D, and thus a unique Nash equilibrium. The other five games have multiple Nash equilibria and no dominant strategies. In Battle of the Sexes, the strategies are strategic substitutes (a player's strategy becomes less attractive, the more likely the opposing player is to choose it), so that the two pure-strategy Nash equilibria are asymmetric; in the other four games with multiple equilibria, the strategies are strategic complements (a strategy becomes more attractive as the likelihood of the opponent playing it increases), so that their pure-strategy Nash equilibria are symmetric. For ease of exposition, we have ordered the two strategies in each game in such a way that the first strategy is "nice" in the sense that it tends to be associated with higher payoffs for the other player than the second strategy; similarly, we have labeled the actions

C (for "cooperate") and D (for "defect"), though these terms are more literally meaningful in some games (such as Prisoners' Dilemma) than in others.

Because each game has only two actions, a player's strategy can be characterized by the associated probability of choosing C. Thus, any *strategy pair* (one strategy for each player) can be written in the form (Prob(Row player chooses C), Prob(Column player chooses C)). The standard game-theoretic prediction for these games is *Nash equilibrium*: a strategy pair in which both players' simultaneously maximize their own payoff, given the other player's strategy. Table 1 shows the games' Nash equilibria, along with the associated probabilities of C choices and action changes from one round to the next; these probabilities will serve as useful benchmarks for our results, even though we are not explicitly testing the equilibrium predictions in this paper. (Note that uniform random play—the (0.5, 0.5) strategy pair—is not a Nash equilibrium of any of these games.)

Table 1: Characteristics of Nash equilibrium play

Game	Strategy pair	Prob(C choice)	Prob(action change)	Game	Strategy pair	Prob(C choice)	Prob(action change)
CG	(1, 1)	1.000	0.000	PD	(0, 0)	0.000	0.000
	(0, 0)	0.000	0.000				
	(1/3, 1/3)	0.333	0.444				
BoS	(1,0)	0.500	0.000	SHH, SHM, SHL	(1, 1)	1.000	0.000
	(0,1)	0.500	0.000		(0, 0)	0.000	0.000
	(3/8, 3/8)	0.375	0.469		(2/3, 2/3)	0.667	0.444

Versions of these games have been used in many experiments. We do not attempt a review of all relevant literature here; much can be found in the sections on coordination games and social dilemmas in Camerer (2003) and Kagel and Roth (1995).

2.1 Experimental design and related literature

As already mentioned, our primary design variable is the matching treatment. In the *fixed-pairs* (F) treatment, subjects played all rounds of a game against the same opponent, though opponents did change from game to game. In the *random-matching* (R) treatment, subjects were randomly assigned to opponents in each round, with every potential opponent (including the previous-round opponent) equally likely. Since the number of rounds of each game was more than the number of potential opponents in an experimental session, subjects in the R treatment faced the same opponent more than once in the same game; however, as no identifying information about opponents was given to subjects, they would never have been able to tell whether—and in which round—they had previously been matched to their current opponent.

In contrast to many experiments, we use a *limited-information* design, under which there is no display of the game's payoffs to subjects—either publicly or privately. Instead, subjects receive some payoff information as part of their end-of-round feedback; specifically, they are informed after each round of their opponent's choice and their own payoff in the just-completed round. While this is enough information to allow subjects to piece together the relationship between outcomes and their own payoffs within a few rounds, it differs from the usual complete-information treatment in two notable ways. First, subjects never receive information about their opponents' payoffs; we believe it is reasonable to expect that the lack of this particular information should serve to undermine any effects on behavior of other-regarding preferences, even in games such as PD where one might normally expect such preferences to be present. Second, it is exceedingly unlikely that the structure of the game was common knowledge amongst the subjects.

There has been some previous work examining the effects of matching mechanisms. Much of this work has involved social dilemmas such as Prisoners' Dilemma and public-good games, played under complete information. These experiments typically did not look for effects on learning, but rather for effects on cooperative behavior; under complete information, one might expect to see more cooperation under fixed pairs—where incentives for reputation building are stronger—than under random matching, and that this effect would be stronger in early rounds than later ones (as the value of a reputation should decline as the number of remaining rounds of play becomes small). However, Andreoni and Croson (2008), summarizing the early literature on comparisons of fixed pairs versus random matching in public-good games, found no consistent relationship between the two: in some experiments, contributions were higher under fixed pairs, while in others, they were higher under random matching. Studies of the Prisoners' Dilemma, on the other hand, have typically found higher levels of cooperation under fixed pairs (Ahn et al., 2001; Duffy & Ochs, 2008), and similar results have been found in other environments that have features of social dilemmas (Abbink, 2004; Charness & Garoupa, 2001; Huck et al., 2001). Mixed results were also found in coordination games whose pure-strategy Nash equilibria can be Pareto-ranked; Van Huyck et al. (1990) and Clark and Sefton (2001) found higher efficiency under fixed pairs than random matching, while Schmidt et al. (2003) found no difference in aggregate play between these mechanisms.

We stress that in all of the studies mentioned above, subjects were given complete information about the games' payoffs. It is thus difficult to disentangle differences in learning between treatments from differences in attempts at reputation building, or in the prevalence of early-round signaling. There have been far fewer experiments looking at different matching mechanisms under limited information. One such experiment was that of Chen (2003), who compared two pricing mechanisms for allocating shared resources within an organization. Subjects made decisions under one of these mechanisms, under either fixed pairs or random matching, and under either synchronous or asynchronous decisions. Subjects were given only "extremely limited" information; they were not informed of the structure of the game before it began, and at the end of each round, they were told their own action and own payoff, but received no feedback at all about opponent actions or payoffs. (There was also one treatment with complete information.) Chen's goal was to compare the pricing mechanisms, not the matching treatments, but examination of her results suggests that there were no substantial or significant differences between the matching mechanisms.

2.2 Experimental procedures

Subjects in our experiment played all six games under either fixed pairs or random matching. We used a two-population design, with half of the subjects in each session designated "row players" and half "column players". Subjects of one type were only matched to subjects of the other type. The ordering of the games was always CG-SH-BOS-SH-PD-SH, but we varied the order in which the three Stag Hunt games were played, in an attempt to control for any sensitivity of behavior to the order in which the games are played. The orderings we used were SHH-SHM-SHL, SHM-SHL-SHH, and SHL-SHH-SHM. We also varied the order in which the actions appeared: C-D (cooperative action on top or left) and D-C (cooperative action on bottom or right). Our manipulations of game ordering, action ordering, and matching mechanism were between-subjects, while our manipulation of the game was within-subject.

Experimental sessions took place at the Kyoto Experimental Economics Laboratory (KEEL) at Kyoto Sangyo University. Subjects were primarily undergraduate students, recruited via a database of participants in other experiments and via advertisements posted on campus. No subject participated in more than one session. At the beginning of a session, each subject was seated at a computer and given a set of written instructions (an English translation, and the raw data from the experiment, are available from the corresponding author upon request). After a few minutes, the written instructions were read aloud by the monitor in an effort to make the rules common knowledge. All subjects were seated in the same room, but partitions prevented them from seeing others' computer screens, and subjects were asked not to communicate with each other during the session. The experiment was programmed in the Japanese version of the z-Tree experimental software package (Fischbacher, 2007), and all interaction took place via the computer network. Subjects were asked not to write down any results or other information.

At the beginning of an experimental session, subjects were told their type (row or column player), and how many rounds of each game they would be playing. Prior to the first round of each game, they were reminded that they were beginning a new game. In every round, subjects were prompted to choose one of the two possible actions, which were given generic names (R1 and R2 for row players, C1 and C2 for column players). After all subjects had made their choices, each was told her own action choice, her opponent's action choice, and her own payoff. In the R treatment, subjects were matched to a new opponent after each round, whereas in the F treatment, subjects were matched to a new opponent when the 40 rounds of one game ended and they began a new game; in both treatments, subjects were informed of the matching mechanism in both oral and written instructions.

At the end of an experimental session, one round of one game was randomly chosen, and each subject was paid 200 yen (at the time of the experiment, equivalent to roughly $1.90) for each point earned in that round. In addition, subjects were paid a showup fee of 3000 yen, from which negative payoffs were subtracted, if necessary.

3. Experimental results

A total of 13 sessions were conducted: 6 of the R treatment and 7 of the F treatment. The number of subjects varied from 6 to 28 in the F sessions, and from 10 to 26 in the R sessions.

3.1 Aggregate behavior

We begin by providing summary statistics of subject behavior in our experiment. Table 2 reports the frequencies of C choices and action changes from the previous round. The first of these variables will be used as a measure of how "good" outcomes are in a particular game, treatment, and round, while the second will be used as a measure of the extent to which individual-subject behavior has converged to pure-strategy play. Since we are primarily interested in how play changes over time, we disaggregate the data by blocks of rounds. For each game, one row of the table shows the data from rounds 1-5 of each game, as a proxy for initial behavior. A second row shows the corresponding data from rounds 36-40 of each game, which we use as a proxy for endgame behavior, and a third row shows the corresponding data over all rounds. In addition to the levels of these statistics, the table shows the results of robust rank-order tests of significance of differences between the F and R treatments for each game, criterion, and time period, using session-level data (see Siegel and Castellan, 1988, for descriptions of the nonparametric tests used in this paper).

In both treatments, frequencies of C choices begin at levels comparable to those implied by uniform random play, as one would expect since subjects initially have no information about payoffs. Indeed, for rounds 1-5, nonparametric Wilcoxon signed-ranks tests find significant differences from uniform random play (two-tailed test, p-values of 0.05 or lower) in only two of the six games in the F treatment and none in the R treatment, and we find no significant differences between the treatments. In later rounds, on the other hand, we find many differences, and when significant, they always point in the same direction: more C choices, and fewer action changes, in the F treatment than in the R treatment. These differences are usually consistent with better outcomes, and quicker convergence, under fixed-pairs matching; the lone exception is Prisoners' Dilemma, where higher levels of C choices are associated with slower convergence to the unique Nash equilibrium (D,D).

Table 2: Observed frequencies from experiment

Game	Rounds	C choices		Action changes	
		F	R	F	R
CG	1-5	0.533	0.530	0.463	0.467
	36-40	0.789	0.891	0.183	0.156
	All	0.684	0.750	0.261	0.269
BoS	1-5	0.454	0.463	0.417	0.469
	36-40	0.494**	0.441	0.105	0.161
	All	0.466***	0.431	0.171***	0.268
PD	1-5	0.454	0.424	0.273	0.322
	36-40	0.206*	0.041	0.057	0.037
	All	0.290**	0.126	0.122	0.123
SHH	1-5	0.454	0.478	0.295	0.274
	36-40	0.392***	0.085	0.060	0.091
	All	0.404***	0.203	0.108	0.147
SHM	1-5	0.421	0.483	0.344	0.406
	36-40	0.457**	0.033	0.030	0.041
	All	0.446**	0.125	0.093*	0.117
SHL	1-5	0.519	0.511	0.537	0.515
	36-40	0.784	0.461	0.051	0.096
	All	0.746	0.499	0.150**	0.233

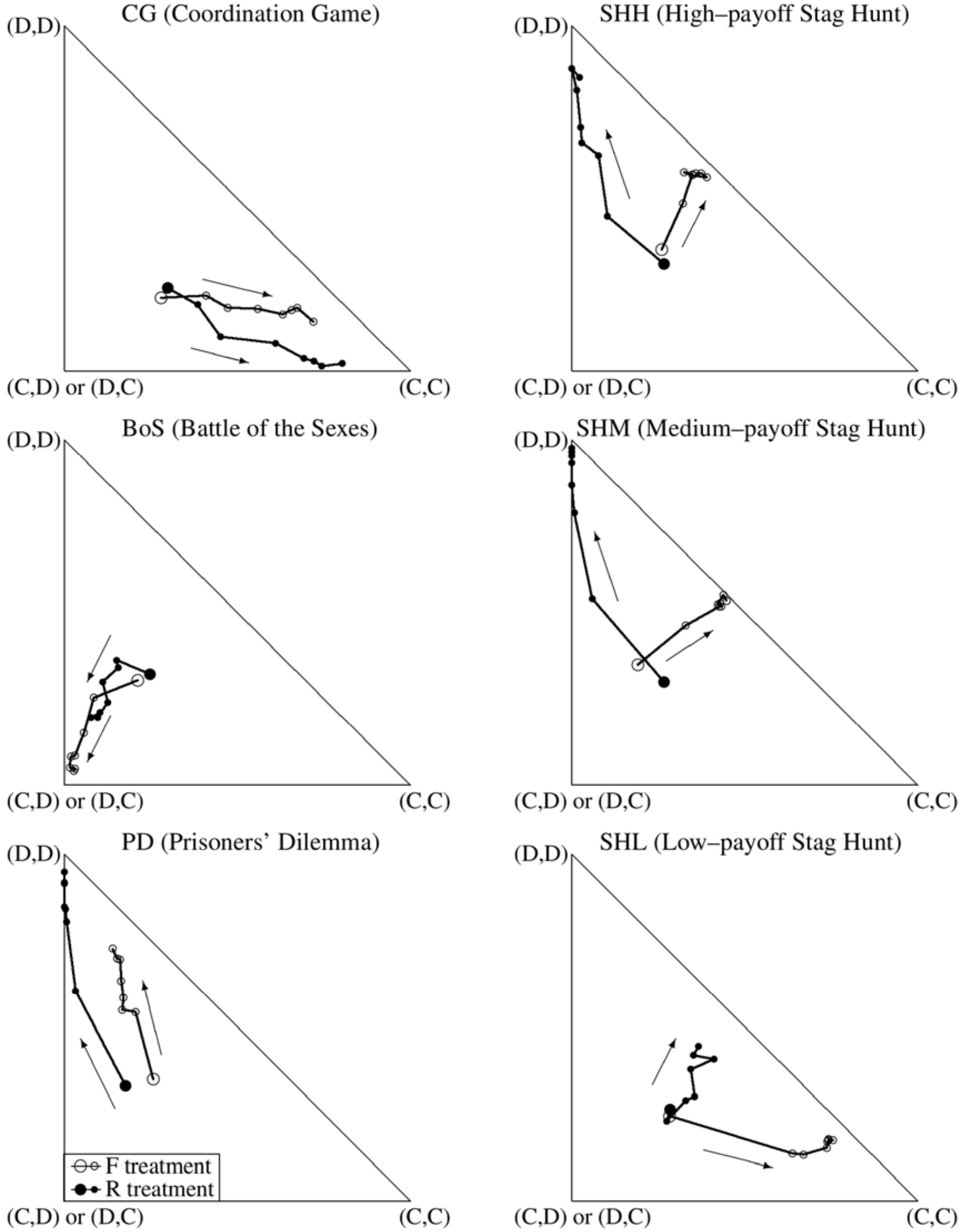
** (**,***): For given game, criterion, and rounds, the value of this statistic is significantly different from the corresponding R treatment statistic at the 10% (5%, 1%) level (two-tailed robust rank-order test, session-level data, pooled action and game orderings).*

3.2 Behavior dynamics

In Figures 2 and 3, we take a closer look at how behavior changes over time. Figure 2 shows the relative frequency of each of the three types of outcome (both players choose C, both choose D, or exactly one chooses C) in each five-round block of each game, under each treatment.

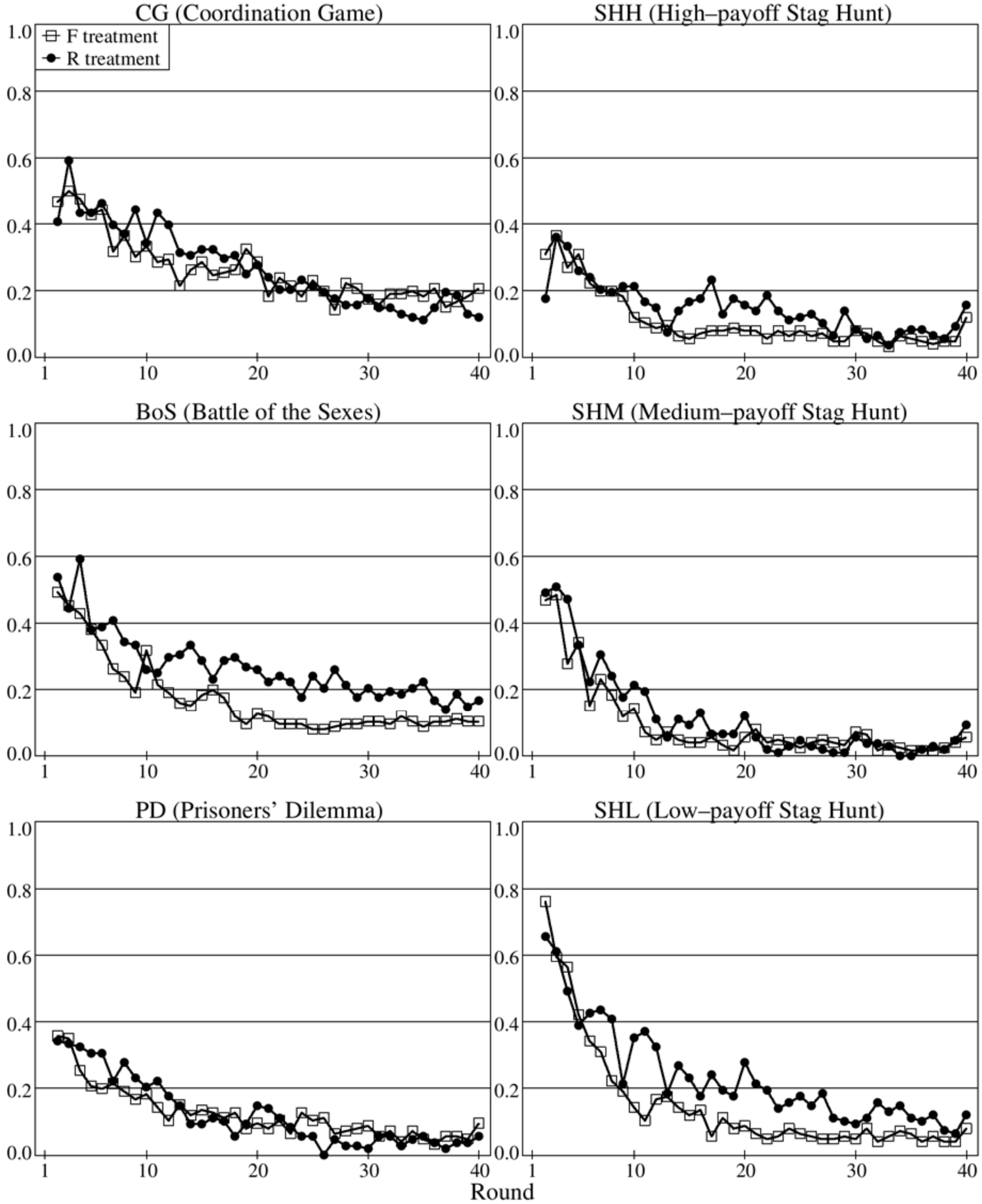
Figure 2: Time paths of experiment aggregate outcome frequencies (five–round blocks)

Large (small) circles represent averages of rounds 1–5 (6–10, 11–15, etc.); arrows indicate time trend.



Arrows indicate the direction of motion of the time paths. As the figure shows, the dynamics of aggregate outcome frequencies vary substantially across games and between F and R treatments. In all games and both treatments, these frequencies begin near the point (1/4, 1/4) implied by uniform random play, then move in the direction of one of the pure-strategy pairs. In CG, both time paths move in the general direction of the Pareto efficient (C,C) Nash equilibrium, but the path for the R treatment ultimately gets closer, suggesting that some pairs in the F treatment become "stuck" at the Pareto inefficient (D,D) equilibrium.[2] In BoS, convergence toward the Nash equilibrium (C,D) and (D,C) pairs is faster and more uniform in the F treatment, while in PD, convergence toward the Pareto dominated Nash equilibrium (D,D) is faster in the R treatment.

Figure 3: Frequency of action changes from previous round in experiment



In the three Stag Hunt games, qualitative differences between the two matching mechanisms are more pronounced. In SHH and SHM, time paths under random matching converge almost completely to the Pareto inefficient (D,D) outcome, while under fixed pairs, play converges

[2] Of the 126 pairs of subjects in the F treatment, ten pairs play the (D,D) pair in each of the last 5 rounds, while two others do so in 4 of those rounds, suggesting that among a small but non-negligible number of pairs, behavior has converged to the (D,D) outcome.

roughly to equal frequencies of the (D,D) and the Pareto efficient (C,C) outcomes; in both treatments, miscoordination on the (C,D) and (D,C) pairs has nearly died out by the end of the game. In SHL, the time path for the F treatment moves in the direction of (C,C), but a few pairs get stuck at (D,D) instead. The path for the R treatment shows some tendency toward roughly equal frequencies of (D,D) and (C,C) outcomes (that is, some *sessions* get stuck at the inefficient (D,D) outcome), but the frequency of miscoordination stays comparatively high.

Figure 3 shows the relative frequency of action changes in each round. Consistent with Table 2, subjects in all games and both treatments change actions with a frequency that is initially relatively high, but decreases throughout the game as subjects learn about the payoff structure and how others play. In some games, there is little apparent difference between treatments, but when a difference is seen (for example, in BoS, SHH, and SHL), we see fewer strategy changes in the F treatment than in the R treatment, suggesting that in the former, subjects are quicker to settle upon one of the action choices.

By and large, therefore, these figures confirm that the manner in which behavior changes over time is often sensitive to the matching mechanism. Individual subjects' decisions converge toward pure-strategy play, as evidenced by the steadily decreasing frequency of action changes. In some games, the rate of decline is similar in both treatments, while in others, it is faster in the F treatment. In five of the six games, play either converges to a better aggregate outcome (the three Stag Hunt games), converges more quickly to a good outcome (Battle of the Sexes), or converges more slowly to a bad outcome (Prisoners' Dilemma) under fixed pairs than under random matching. In the remaining game (Coordination Game), average behavior tends toward a somewhat worse outcome under fixed pairs, as a nontrivial fraction of pairs get stuck at the lower-payoff equilibrium, as compared to random matching, where all sessions converge to near-complete play of the higher-payoff equilibrium.

3.3 Parametric statistics

In this section, we report the results and implications of several regressions. This gives us the opportunity not only to assess the significance of the suggestive results seen in the previous sections, but also to increase the power of our hypothesis tests by using the entire data set for each game rather than limiting ourselves to data from individual treatments of each game.

We run two sets of probits with individual-subject random effects. In the first, the dependent variable is an indicator for a C action choice, while in the second, the dependent variable is an indicator for a change in action choice from the previous round. In order to determine the effect of the matching mechanism, we use as our primary explanatory variables an indicator for the F treatment and its product with the round number (so that the baseline is the R treatment). As controls, we include a variable for the round number itself and indicators for two of the three game orderings (M-L-H and L-H-M) and one of the two action orderings (D-C). We estimate coefficients separately for each regression, using Stata (v. 10).

Table 3: Probit coefficients, standard errors, and p-values for significance of coefficients

	CG		SHH	
Variable(s)	C choices	Action changes	C choices	Action changes
F indicator	0.080	−0.203**	−0.129	−0.089
	(0.114)	(0.098)	(0.214)	(0.118)
F*round	−0.013***	0.008***	0.047***	−0.016***
	(0.003)	(0.003)	(0.004)	(0.004)
Joint signif.	p<0.001	p≈0.019	p<0.001	p<0.001
	BoS		SHM	
F indicator	−0.071	−0.235***	−0.046	−0.242**
	(0.195)	(0.112)	(0.197)	(0.106)
F*round	0.006**	−0.016***	0.084***	0.002
	(0.003)	(0.003)	(0.004)	(0.005)
Joint signif.	p≈0.082	p<0.001	p<0.001	p≈0.039
	PD		SHL	
F indicator	0.323*	−0.299***	0.388***	−0.008
	(0.177)	(0.110)	(0.149)	(0.097)
F*round	0.023***	0.016***	0.033***	−0.023***
	(0.004)	(0.004)	(0.003)	(0.004)
Joint signif.	p<0.001	p<0.001	p<0.001	p<0.001

** (**,***): Coefficient is significantly different from zero at the 10% (5%, 1%) level.*

For brevity, we do not report all coefficient estimates here (the full results are available from the corresponding author upon request). Rather, we concentrate on the two F-treatment variables—their coefficient estimates, standard errors, and significance—as well as p-values associated with tests of their joint significance. Table 3 shows these results, and provides strong evidence that the matching mechanism has an effect on learning. In all eighteen of the probits, at least one of the F-treatment variables is significant at the 5% level or better. Also, the two variables are always jointly significant at the 10% level, and most pairs are significant even at the 1% level.

We next concentrate specifically on the incremental effect of switching from the R to the F treatment: the sum of the shift effect (based on the variable $\beta_{\text{F treatment}}$) and the time-dependent effect (based on the expression $\beta_{\text{F*round}} \cdot t$), where β_{x} is the coefficient of the variable x and t is the round number.

Figure 4: Estimated incremental effect of F (vs. R) treatment on C choices and action changes based on probit results *(circles and squares represent point estimates; line segments represent 95% C.I.)*

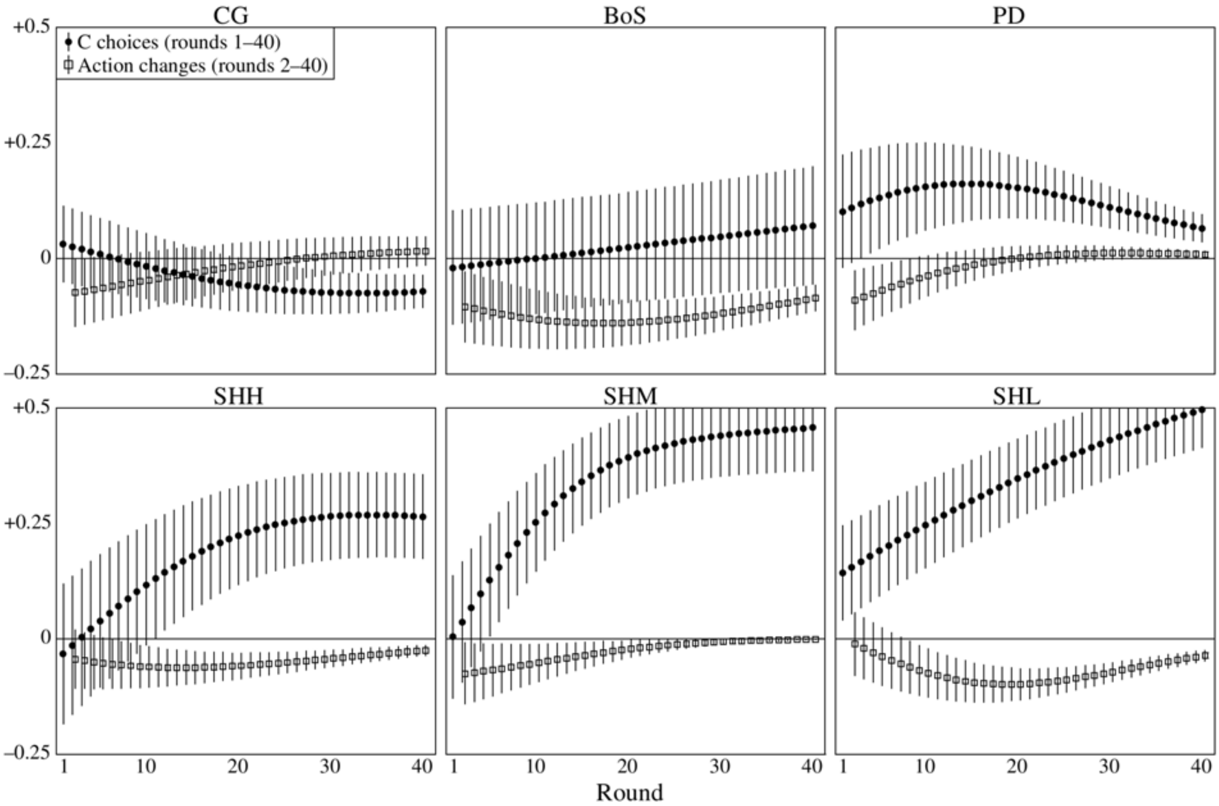


Figure 4 shows these estimated incremental effects on C choices and action changes. The figure shows point estimates (as closed circles or boxes) and 95% confidence intervals (as line segments) for these effects, as well as the horizontal segment corresponding to a zero effect. The figure shows that the effect of the F treatment versus the R treatment on C choices varies across games. In CG, it decreases over time, becoming negative and significant by the second half of the session (reflecting the pairs that get stuck at the (D,D) outcome). In BoS, the effect is never significant, though it is usually positive. In PD and the three Stag Hunt games, the effect becomes positive and significant early on, and remains so throughout. The size of the effect peaks in the middle of the session in PD and near the end of the session in SHH, but continues growing over time in the other two Stag Hunt games. By the last round of SHM and SHL, this incremental effect is almost one-half, which is quite substantial for a variable bounded by 0 and 1.

The effect of the F treatment versus the R treatment on the frequency of action changes is reasonably consistent across games. It is nearly always negative (that is, fewer action changes in the F treatment than in the R treatment)—and usually significantly so—in early rounds, with the magnitude of the effect declining over time, as subjects in both F and R treatments tend toward pure-strategy play.

To summarize, these results replicate the patterns discussed in Section 3.2. Typically, outcomes are better, and learning is faster, in the F treatment than in the R treatment. Differences between the treatments are small in early rounds (before subjects have gotten feedback about the games' payoffs), but become more pronounced over time. In some cases, these differences grow over the course of a game; in others, they shrink in later rounds, as behavior in the R treatment "catches up" to that in the F treatment.

4. Learning model analysis

In this section, we present simulation results based on a model of individual learning, in an attempt to explain the differences we have seen between fixed pairs and random matching. Since our experiment involved limited information, the model we use is adapted from Erev and Barron's (2005) RELACS (reinforcement learning among cognitive strategies) model. The RELACS model has been found to perform well in characterizing behavior in a variety of individual decision tasks under limited information (Erev et al., 2008; Grosskopf et al., 2002; Munichor et al., 2006). Our aims here are more modest; we wish merely to show that there exists a learning model with implications qualitatively similar to our experimental results.

4.1 Description of the model

According to the RELACS model, a player chooses an action in a given round according to a two-step process (see Erev and Barron, 2005, pp. 919-922, for a fuller description). In the first step, the player chooses one of several available decision rules ("cognitive strategies"), and in the second step, follows the chosen decision rule to determine the action that will be played. After the game is played, the feedback received by the player is used to update the likelihood of choosing each cognitive strategy, as well as the action choices themselves.

We use the same three cognitive strategies used by Erev and Barron, and add one additional one to yield our modified RELACS model. According to the *fast best reply* rule, a player chooses the action with the higher "recent payoff" (approximately a weighted average of past payoffs). For an action $j = C, D$, the recent payoff $R_j(t)$ is initially the expected payoff from uniform random choice by both players (e.g., 5 in PD), and is updated (in rounds it is chosen) according to $R_j(t+1) \ = \ (1-\beta)R_j(t)+\beta \ v(t),$ where $v(t)$ is the payoff obtained from action j in round t, and $\beta \in (0,1)$ is a "recency" parameter that captures how recent payoffs are weighted relative to earlier ones. If both actions have the same recent payoff, the player chooses randomly between them.

According to the *slow best reply* rule, a player chooses an action according to a probability distribution. The probability of choosing action j in round t is given by

$$p_j(t) = \frac{\exp\left(\lambda W_j(t)/S(t)\right)}{\exp\left(\lambda W_C(t)/S(t)\right)+\exp\left(\lambda W_D(t)/S(t)\right)},$$

where $W_j(t)$ is a weighted average of past payoffs, $S(t)$ is a measure of payoff variability, and $\lambda \geq 0$ is an index of sensitivity to payoff differences. $W_j(t)$ is similar in nature to the recent payoff measure $R_j(t)$; $W_1(t)$ is the expected payoff from uniform random choice by both players, and $W_j(t)$ is updated in rounds when j is chosen according to the rule $W_j(t+1) = (1-\alpha)W_j(t) + \alpha v(t)$,

where $\alpha \in (0, \beta)$ is another recency parameter. The payoff variability measure $S(t)$ also adjusts over time; $S(1)$ is the expected absolute difference of payoffs from the payoff implied by random choice (e.g., (|7-5|+|1-5|+|8-5|+|4-5|)/4 = 2.5 in PD), and

$$S(t+1) = (1-\alpha)S(t) + |\, v(t) - \text{Max}\{Last_C, Last_D\}\,|,$$

where $Last_C$ and $Last_D$ are the most recent payoffs from the two actions.

According to the *case-based reasoning with loss aversion* rule, a player samples from her past results, and chooses the action that (based on the sample chosen) resulted in higher payoffs, subject to not incurring more and deeper losses. Specifically, a two-stage process is involved. In the first stage, one previous result is randomly drawn for each action. If the actions led to the same payoffs, the player draws again for each action, and continues this random sampling until the tie is broken. (If the player has never played one of the actions, or if all previous outcomes yielded the same payoffs, the process ends with the player randomly choosing one of the actions.) In the second stage, the player samples an additional k previous results for each action, and determines (i) the number of times a negative payoff (loss) was incurred, and (ii) the sum of losses incurred, in those k+1 outcomes observed for each action. The player then chooses the action that had the higher payoff in the first stage, unless it performed worse in both part (i) and part (ii) of the second stage, in which case the other action is chosen.

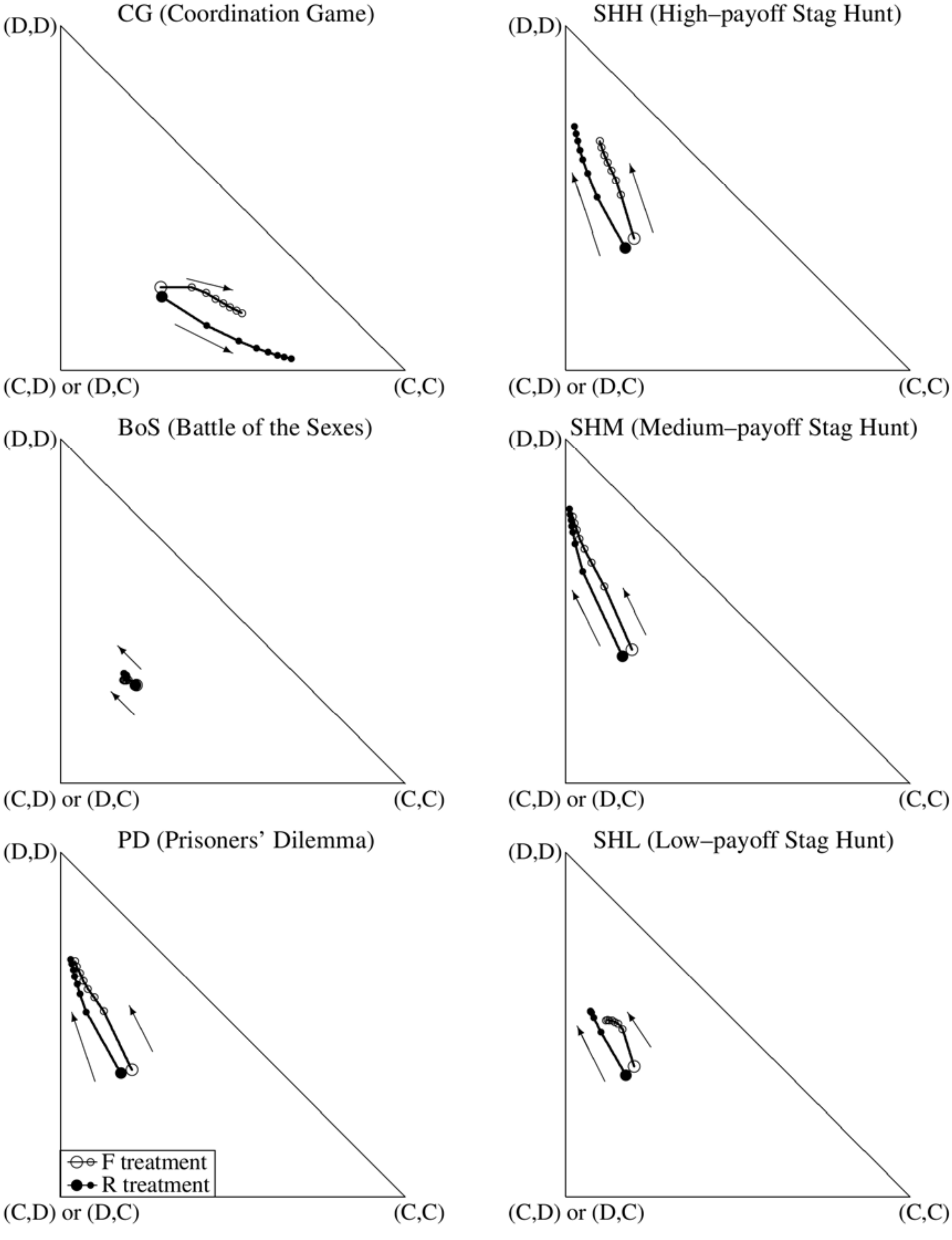
The new cognitive strategy we consider is the *tit-for-tat* rule, according to which the player chooses the same action as her previous-round opponent did (in the first round, the player randomizes). We add this cognitive strategy for two reasons: (1) to reflect the fact that subjects in our experiment are playing games (and are aware of this fact), in contrast with the single-person decision tasks faced by subjects in the situations studied by Erev and Barron; and (2) to account for previous research showing that tit-for-tat is both a commonly-played and a successful strategy in repeated versions of Prisoners' Dilemma (Axelrod, 1984). While our limited-information setup makes it less likely that subjects in the experiment recognized the structure of our PD game, adding tit-for-tat as a cognitive strategy allows us to observe whether it is chosen by simulated subjects.

Finally, the cognitive strategies themselves are chosen in each round according to the same slow best reply rule that formed the second cognitive strategy above.

4.2 Simulation design and results

Figure 5: Time paths of simulation aggregate outcome frequencies (five–round blocks)

Large (small) circles represent averages of rounds 1–5 (6–10, 11–15, etc.); arrows indicate time trend.



The simulations were programmed in GAUSS; the programs are available from the corresponding author upon request. Following Erev and Barron (2005), we use the following model parameters: α=0.00125, β=0.2, λ=8, and k=4. For each game and matching mechanism, we run 10,000 40-round simulated experimental sessions. For the random-matching treatment, each session is made up of 20 automated subjects, randomly paired in each round. For the fixed-pairs treatment, each session is again made up of 20 subjects, but they are randomly paired only at the beginning of the first round, and remain paired to the same opponent throughout all 40 rounds of the game.

Figure 5 shows the relative frequency of each type of outcome (both choose C, both choose D, or exactly one chooses C) in the simulations, similarly to Figure 2 for the experiment. By contrast with Figure 2, the trajectories in this figure primarily show differences in speed of convergence, rather than differences in the likely asymptote. However, the differences between treatments are qualitatively similar to those observed in Figure 2. In CG, convergence to the (C,C) outcome is faster in the R treatment than in the F treatment, as some pairs in the latter get "stuck" on the Pareto inefficient (D,D) outcome. In BoS, differences are quite small and difficult to see, but the F treatment simulations show slightly more movement toward (C,D), and less play of (C,C) and (D,D), than the R treatment. In PD, both treatments show movement toward the unique equilibrium (D,D), but this movement is faster in the R treatment. In all three of the Stag Hunt games, the simulations tend toward the Pareto inefficient (D,D) outcome, but as in PD, this tendency is always somewhat stronger in the R treatment.

Thus, the results of our simulations show that although the version of RELACS we use does not perfectly characterize behavior in the experiment, it does pick up our main result: the qualitative differences between fixed pairs and random matching.[3] This is significant, since it suggests that the model might be useful for forming predictions about the effects of fixed pairs versus random matching in additional games. In particular, the model provides an explanation for the seemingly counter-intuitive PD result of slower convergence in the F treatment than in the R treatment. To see how this can happen, consider a population using only the fast best reply cognitive strategy. Initially, the recent payoff for both C and D actions is 5, so in the first round, each action is chosen with probability one-half. If the outcome of the first round is (C,C)—which happens with probability one-fourth—then both players in the pair earn a payoff of 7, which increases the recent payoff to C, while the recent payoff to D remains at 5. If the outcome is (D,D)—which also happens with probability one-fourth—the payoff of 4 means that the recent payoff to D falls while the recent payoff to C stays constant. In either of these cases, under fixed pairs both players will choose C in all rounds thereafter. In the other two cases, (C,D) and (D,C), both players would choose D from the second round onwards. There is thus a one-half probability under fixed pairs that a pair gets "stuck" on the good (C,C) outcome. Under random matching, by

[3] One can conjecture several reasons for the inability of this version of the model to characterize the data even better, but we should point out that our implementation of RELACS did not give it its best chance. Rather than fitting model parameters to the data, or looking for reasonable ones using a grid search, we limited our analysis to the parameters used by Erev and Barron. While their parameters were found to work well for individual decision tasks, which by their nature are stationary, it may be that strategic games, which are non-stationary due to the effect of changing opponent behavior, require different parameters. It may also be that subjects simply learn in a different manner when they know they are playing a game, in which case adding or removing cognitive strategies—along the lines of our addition of tit-for-tat—might be warranted.

contrast, a player whose first-round outcome was (C,C)—and hence chooses C in the second round—has about a one-half chance of meeting someone who plays D in the second round, rather than the zero chance of this under fixed pairs. In that case, the player's recent payoff to C would fall below 5 (the recent payoff to D), so that she would choose D in the next round. Since D strictly dominates C, over time the recent payoff to D will typically be higher than that to C, so eventually C will die out under random matching. (Indeed, simulations of a variant of the RELACS model with fast best reply as the only available cognitive strategy yield C choices persisting with probability one-half under fixed pairs, but dying out under random matching.) To the extent that individuals have other cognitive strategies available besides fast best reply, the differences between F and R treatments will be less stark, but the direction of the effect will be the same: faster convergence to the (D,D) outcome under random matching than under fixed pairs.

Table 4: Mean cognitive strategy probabilities, round 1000 of simulations

Game	Random matching				Fixed pairs			
	Fast BR	Case-based	Slow BR	TfT	Fast BR	Case-based	Slow BR	TfT
CG	0.571	0.145	0.027	0.257	0.579	0.133	0.021	0.267
BoS	0.915	0.001	0.084	0.000	0.907	0.016	0.076	0.000
PD	0.853	0.080	0.030	0.038	0.769	0.138	0.047	0.046
SHH	0.788	0.049	0.045	0.118	0.651	0.145	0.040	0.164
SHM	0.464	0.408	0.002	0.125	0.515	0.365	0.001	0.119
SHL	0.249	0.259	0.266	0.226	0.249	0.260	0.267	0.223

In order to further understand the behavior of our modified RELACS model, we look at how cognitive strategies evolve over time. Initially, each of the four is chosen with equal probability, but over time the probability of a given cognitive strategy reflects its success. Over only 40 rounds, there is little opportunity for these probabilities to change by large amounts—over all games and both treatments, they range only from 0.224 to 0.279 in the 40th round. We therefore extend the simulations to 1000 rounds, and look at the cognitive strategy probabilities after the 1000th round. The results are shown in Table 4. In all games except SHL (where the predominance of negative payoffs means that most reinforcement is negative, so that chosen strategies tend to become less frequent), there are substantial differences across cognitive strategies by round 1000. Fast best reply is by far the most successful cognitive strategy: it is chosen with a plurality of the probability in all games but SHL. Case-based reasoning does relatively well in the two games with losses possible (as it aids coordination when the opponent is using the same cognitive strategy). Slow best reply does badly in all games except SHL, likely because it causes the less successful action to be chosen with substantial probability (it would likely have fared better in a game with strictly opposed preferences, such as matching pennies, where less historical success might indicate better present success). Tit-for-tat is most successful in the games with strategic substitutes, where it is equivalent to best reply to the previous action; in BoS and (interestingly, given the previous literature) PD, it does badly.

There are no striking differences in cognitive strategy probabilities between the F and R treatments—even after 1000 rounds—which partly explains why the only differences observed in Figure 5 were in speed of convergence. The largest differences we do see occur in PD and SHH, where fast best reply is less successful under fixed pairs than under random matching, with case-based reasoning correspondingly more successful. We surmise that this effect is due to the main difference between the two cognitive strategies in these games (with negative payoffs impossible): case-based reasoning chooses the action that earned a higher payoff in one random draw, while fast best reply chooses the action that has earned a higher weighted average payoff over all rounds. Ordinarily, it would seem desirable to use the latter cognitive strategy, since it is less sensitive to outliers. This is indeed what we see under random matching, but under fixed pairs, both cognitive strategies are more likely to yield the same choice (since there is less variation in choices within an individual than across individuals within a session), so that fast best reply has more difficulty in driving case-based reasoning out of the population of cognitive strategies. The implication of this reasoning is that even though we continue observing differences between F and R treatments in the probabilities of cognitive strategies, the probabilities of the actions themselves differ much less between these treatments—exactly what was seen in Figure 5.

5. A new experiment

As a robustness test for both our experimental results and our simulation predictions, we conducted a follow-up experiment using another game. The game is chosen from the three games used in Battalio et al.'s (2001) experiment (see Figure 6). These games are most similar to our coordination game (CG) and our high-payoff stag hunt game (SHH), in that strategies are strategic complements and payoffs are non-negative. However, unlike CG, players' interests are not perfectly aligned in these new games, and unlike SHH, neither strategy is risk-free. All three games have the same set of Nash equilibria: (C,C), (D,D), and a mixed-strategy equilibrium in which each player chooses C with probability 0.8.

Figure 6: Games used by Battalio et al. (2001)

		Player 2 C	Player 2 D
Player 1	C	45,45	0,42
	D	42,0	12,12

0.6ρ game

		Player 2 C	Player 2 D
Player 1	C	45,45	0,40
	D	40,0	20,20

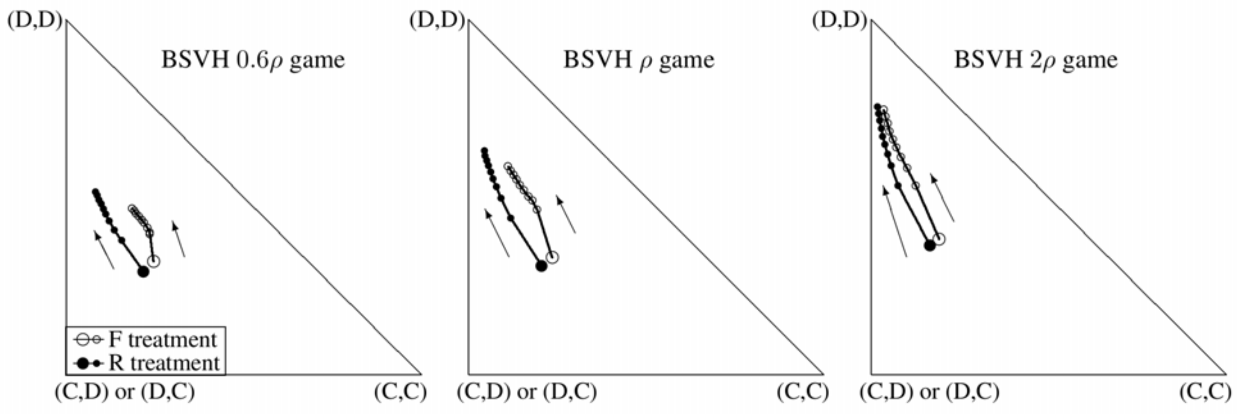
ρ game

		Player 2 C	Player 2 D
Player 1	C	45,45	0,35
	D	35,0	40,40

2ρ game

In order to form predictions for the experiment, we simulate subject behavior using the modified RELACS model from Section 4. The simulation design was broadly the same as before. We used the same parameters (α=0.00125, β=0.2, λ=8, and k=4), and the same number of simulated sessions of each game and treatment (10,000). The only changes from the earlier simulations are: (1) we ran our new simulations for 50 rounds instead of 40, and (2) we used one-population matching for the random-matching treatment. (Both of these changes reflect procedures we follow in our new experiment, as described in the next section.) Average outcome frequencies from the simulations are shown in Figure 7. In all three games, there is more tendency toward the Pareto inefficient (D,D) equilibrium under random matching than under fixed pairs, but the size of this difference varies across the games: it is largest in the 0.6ρ game and smallest in the 2ρ game. We therefore use the 0.6ρ game in our new experiment.

Figure 7: Battalio et al.'s games: time paths of simulation aggregate outcome frequencies (five–round blocks)
Large (small) circles represent averages of rounds 1–5 (6–10, 11–15, etc.); arrows indicate time trend.



5.1 Experimental design and procedures

In the new experiment, we kept the main features of our previous experimental design, but changed some of the experimental procedures. As before, we varied the matching mechanism (fixed pairs or random matching) and the order in which the actions were presented (C-D or D-C). Subjects played only one game, the 0.6ρ game. Following Battalio et al. (2001), our random-matching treatment used a one-population mechanism (all other subjects in the session were equally likely to be the opponent), rather than the two-population mechanism from our original experiment. As in our previous experiment, but unlike Battalio et al., subjects play under limited information; while they are told that they are playing a game, and how the matching is done, they are never shown a payoff matrix, nor are they ever given feedback about opponent action choices or payoffs. Thus, subjects in this new experiment receive even less information than those in our original experiment, who were informed of opponent action choices. (The implementation of limited information used here is thus very similar to that of Chen (2003)).

The experiment took place at the Scottish Experimental Economics Laboratory (SEEL) at University of Aberdeen, and was programmed in z-Tree (Fischbacher, 2007). Subjects were primarily undergraduate students. None took part more than once in this new experiment, nor did anyone participate who participated in our original experiment. At the beginning of a session, subjects were seated and given written instructions, which were read aloud before play began. These instructions, as well as the z-Tree programs and the raw data, are available from the corresponding author upon request. After any questions were answered, the first round began. As in the original experiment, partitions prevented viewing of others' computer screens, and subjects were asked not to communicate with each other. Each round consisted of simultaneous action choices, then feedback consisting of own choice, own payoff, and total earnings up to that round. After the first round, a subject's history of past feedback was available on her computer screen.

After subjects played fifty rounds, they were paid their total earnings, converted to GBP at an exchange rate of £1 (approximately $1.60 at the time of the experiment) per 100 points. There was no showup fee. Sessions lasted 45-75 minutes, and average payments were roughly £15.

5.2 Experimental results

We conducted 8 sessions (4 each of the F and R treatments) of the new experiment, with a total of 108 subjects. Table 5 shows aggregate frequencies of C choices and action changes for the first five, last five, and all rounds of both treatments. There is little apparent difference between the two treatments in the first five rounds, which is unsurprising since our limited-information design ought to lead to nearly uniform random initial play. (Wilcoxon tests confirm that frequencies of C choices in rounds 1-5 are not significantly different from uniform in either treatment, or even for both treatments pooled together; $p > 0.1$ in all cases.) By the last five rounds, however, striking differences emerge: C choices are much more likely, and action changes much less likely, under fixed pairs than under random matching. These differences are significant at the 5% level, except for action changes in the last 5 rounds, where the difference is only significant at the 10% level. The picture is nearly identical for the aggregate data: significantly more C choices, and fewer action changes, in the F treatment compared to the R treatment. The differences seen here are similar in nature to some of the results in our original experiment.

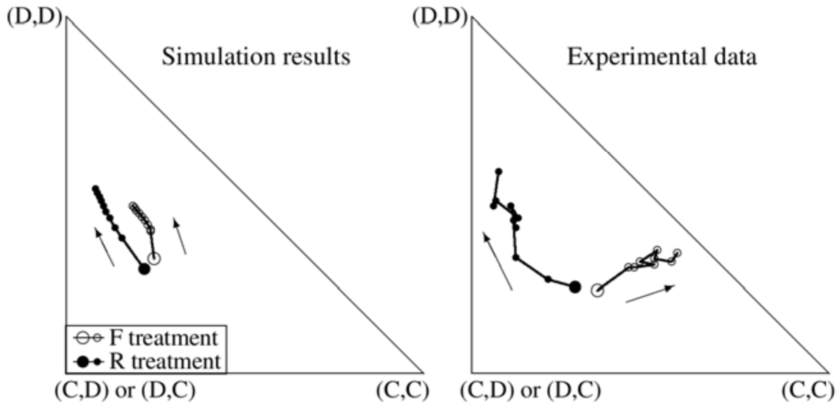
Table 5: Observed frequencies from new experiment (0.6ρ game)

Rounds	C choices		Action changes	
	F	R	F	R
1-5	0.560	0.524	0.532	0.569
46-50	0.624**	0.255	0.148*	0.317
All	0.595**	0.362	0.236**	0.367

** (**, ***): For given game, criterion, and rounds, the value of this statistic is significantly different from the corresponding R treatment statistic at the 10% (5%, 1%) level (two-tailed robust rank-order test, session-level data, pooled action and game orderings).*

We next compare the experimental data to the results from the RELACS learning model simulations in Figure 8. This figure shows the frequency of the possible outcomes in both simulation and experiment, for both treatments and each five-round block. As in Section 4, we are less interested in the learning model's point predictions than in its prediction about the treatment effect. Once again, the difference between the fixed-pairs and random-matching treatments in the experiment is the same as for the simulations: we see here that both experiment and simulation have more (C,C) outcomes, and fewer (D,D) outcomes, in the F treatment than in the R treatment. We therefore conclude that the modified RELACS model can be useful as a source of predictions about the difference between fixed pairs and random matching, though we make no claim about the usefulness of its point predictions.

Figure 8: 0.6ρ game: time paths of aggregate outcome frequencies in simulations vs. experiment (five–round blocks)
Large (small) circles represent averages of rounds 1–5 (6–10, 11–15, etc.); arrows indicate time trend.



6. Discussion

In any strategic situation that is not so trivial that decision makers immediately figure out which actions to choose, it is important to be able to model the way their decision-making behavior adjusts over time (which we have been calling "learning"). In order to successfully model learning, we need to understand which aspects of the situation influence the way individuals learn. In this paper, we examine the effect on learning of the manner in which players are matched with opponents, using six simple two-player games. Each game is played forty times, under either of two commonly-used matching mechanisms. In *fixed-pairs matching*, a player is matched to the same opponent for all rounds of a game. In *random matching*, a player is rematched after every round. The games are played under *limited information*: subjects are never shown any game's payoff matrix, and while they receive information about their own payoffs in the end-of-round feedback, they never receive any information about opponent payoffs. Besides ameliorating any effects of subjects' other-regarding preferences on behavior, our design serves to isolate the effect of the matching mechanism on learning, by limiting the effects of other factors that affect behavior, such as signaling or reputation building.[4]

In the experiment, we find sizable and systematic differences in behavior between fixed-pairs and random matching. Outcomes are typically better under fixed-pairs matching, as cooperative choices are more likely. However, the opposite can happen (as in our Coordination Game), when some pairs of subjects become "stuck" on an inefficient Nash equilibrium under fixed-pairs matching. We typically find faster convergence to pure-strategy play under fixed-pairs matching (with the exception of Prisoners' Dilemma), in that action changes from one round to the next are less frequent than under random matching. The differences between treatments are often small in early rounds, but grow over time, and are not only visible in summary statistics, but confirmed by both non-parametric and parametric statistical tests. We find similar results in a follow-up experiment, using a different game and with some aspects of the experimental procedures altered as a robustness check. The differences we find between fixed pairs and random matching are consistent with the results of simulations based on a variation of Erev and Barron's (2005) RELACS learning model, modified to allow for tit-for-tat play. Notably, while the original RELACS model had been found by other researchers to explain individual decision-making behavior well, to our knowledge it has never previously been used to characterize behavior in strategic games.

The differences we observe are also consistent with previous work comparing fixed pairs and random matching. Specifically, our finding of better outcomes in Prisoners' Dilemma under fixed pairs is similar to the findings of several other researchers in complete-information social dilemmas (Abbink, 2004; Ahn et al., 2001; Charness & Garoupa, 2001; Duffy & Ochs, 2008; Huck et al.,

[4] Because we did give subjects information about opponent action choices in the end-of-round feedback in our main experiment, we cannot completely rule out the possibility that some of the more sophisticated subjects were able to signal in early rounds via their action choices, and that these signals had an effect on opponent choices and coordination. In our follow-up experiment, by contrast, subjects did not receive this information, so such signaling would have been nearly impossible.

2001), though not Andreoni and Croson (2008), who found no systematic difference. Our finding of better outcomes in Stag Hunt games under fixed pairs is similar to what was observed by Van Huyck et al. (1990) and Clark and Sefton (2001) in coordination games with complete information, but not Schmidt et al. (2003), who did not find such a difference. Unlike Chen's (2003) limited-information experiment, our experiment did yield differences between fixed-pairs and random-matching treatments, though it should be noted that none of our games were closely related to the games she used (see Section 2.1 for details). To sum up, then, when a previous experiment has found a difference between fixed pairs and random matching in some game, we have found a difference in the same direction in our most closely related game (though the converse does not always hold, since not all previous experiments found differences).

Our results can—to a fair degree—be explained by two distinct properties of fixed pairs relative to random matching. First, learning about opponent behavior is a simpler task when there is only one potential opponent (fixed-pairs matching) than when there are multiple potential opponents (random matching). This greater simplicity increases the likelihood of faster convergence to a Nash equilibrium under fixed pairs, as observed in most of the games. Second, groups of size two (fixed pairs) are more likely to show heterogeneity in group averages than larger-sized groups (random matching). When the game has multiple equilibria, this increased heterogeneity can lead to different groups converging to different action pairs, with the implication that some groups become stuck on inefficient equilibria. Indeed, of the games with multiple pure-strategy Nash equilibria, all five saw different pairs in the F treatment converging to different equilibria, but only two saw different sessions in the R treatment reaching different equilibria.

Our limited-information design was intended to minimize the effects of factors other than learning, in contrast to previous tests of matching mechanisms (see Section 2.1) that used complete-information designs. As a result, any differences between our results and those from complete-information experiments are likely to be due to these other factors. Many of our results (e.g., usually faster convergence to pure-strategy play under fixed pairs than under random matching) have counterparts in complete-information experiments. However, our finding that the possibility of pairs becoming stuck on inefficient equilibria can lead to *worse* outcomes under fixed pairs than under random matching is generally not observed under complete information. While researchers have found that behavior in complete-information experiments can also be sensitive to early-round outcomes (Van Huyck et al., 1990), the usual consequence is that some subjects under fixed pairs use early rounds to signal cooperative actions (Clark & Sefton, 2001), improving outcomes relative to random-matching treatments (where incentives for such signaling are much weaker). Under limited payoff information, such signaling is much less likely to work, leading to the possibility of worse outcomes under fixed pairs.

The main puzzle in our results comes from the Prisoners' Dilemma in our original experiment, where there was actually slower convergence to equilibrium under fixed pairs than under random matching. Figure 2 showed a persistent, non-negligible fraction of (C,C) outcomes, which in a complete-information setting would suggest some combination of social preferences or supergame behavior. However, our experiment was designed with limited information in order to avoid these possibilities, so it is unclear why there should be so many cooperative action choices.[5]

A few other remarks are warranted at this point. First, since the evolution of behavior over time is sensitive to the matching mechanism used, we suggest that care should be taken whenever drawing conclusions based on data using only one matching mechanism. In cases where the only effect of the matching mechanism is on the speed of convergence, qualitative conclusions will probably be fairly robust. However, in some games, the choice of matching mechanism actually had an effect on the outcome to which behavior converged. In these cases, especially, conclusions about equilibrium selection based on only one mechanism may be misleading. We are not arguing that all experiments should involve multiple treatments with varying matching mechanisms, as we recognize that experimental sessions are costly, and adding treatments with other matching mechanisms will generally imply a sacrifice of other potential design treatments. However, a possible compromise might be the use of additional pilot sessions involving different matching mechanisms, in order to assess the robustness of results to the mechanism used.

Second, as discussed in the introduction, a successful model of learning ought to be able to predict the differences in play seen in our results. It is not immediately clear what features a learning model must have in order to characterize all of our results, but our simulations using the modified RELACS model suggest that it has some potential. A "horse race" comparison of learning models is beyond the scope of this paper, but the possibility certainly exists that other models (or even a different modification of RELACS) can do an even better job. We would welcome further research that examines the ability of other learning models to predict the differences between fixed pairs and random matching that we observed.

Finally, we acknowledge that we have thus far barely scratched the surface in this area. We consider a study of the two most widely-used matching mechanisms in seven fairly well-known games to be a good start, but there are other matching mechanisms, and countless other games in which the choice of matching mechanism could have an effect. Additionally, many implementations of limited information are possible besides those used here; one possible

[5] One possibility is that some of the subjects either assumed or inferred (based on the feedback they received) that the games were symmetric. In that case, they could then determine their opponents' payoffs once they had worked out their own part of the payoff matrix. If both subjects in a pair did so, then the chance of effects from social preferences or supergame behavior could increase substantially. One bit of evidence that suggests this might have happened is that of the 126 pairs in the F treatment of our original experiment, 18 played the (C,C) outcome in at least 9 of the final 10 rounds; that is, much of the (C,C) play was concentrated in a few pairs. (An alternative possibility arises from the RELACS model; see Section 4.2.)

alternative design would not even inform subjects that they are playing a strategic game. We wish to encourage more work in this direction as well.

References

Abbink, K. (2004). Staff rotation as an anti-corruption policy: an experimental study. *European Journal of Political Economy*, *20*, 887-906. DOI: 10.1016/j.ejpoleco.2003.10.008.

Ahn, T. K., Ostrom, E., Schmidt, D., Shupp, R., & Walker, J. (2001). Cooperation in PD games: fear, greed, and history of play. *Public Choice*, *106*, 137-155. DOI: 10.1023/A:1005219123532.

Andreoni, J., & Croson, R. (2008). Partners versus strangers: random rematching in public goods experiments. In C. Plott, and V.L. Smith (Eds.), *Handbook of Experimental Economics Results, v. 1* (pp. 776-783). New York: Elsevier.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Battalio, R., Samuelson, L. & Van Huyck, J. (2001). Optimization incentives and coordination failure in laboratory stag hunt games. *Econometrica*, *69*, 749-764. DOI: 10.1111/1468-0262.00212.

Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton, NJ: Princeton University Press.

Camerer, C., & Ho, T.-H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, *67*, 827-874. DOI: 10.1111/1468-0262.00054.

Charness, G., & Garoupa, N. (2000). Reputation, honesty, and efficiency with insider information: an experiment. *Journal of Economics and Management Strategy*, *9*, 425-451. DOI: 10.1111/j.1430-9134.2000.00425.x.

Chen, Y. (2003). An experimental study of serial and average cost pricing mechanisms. *Journal of Public Economics*, *87*, 2305-2335. DOI: 10.1016/S0047-2727(02)00017-8.

Clark, K., & Sefton, M. (2001). Repetition and signalling: experimental evidence from games with efficient equilibria. *Economics Letters*, *70*, 357-362. DOI:10.1016/S0165-1765(00)00381-5.

Coricelli, G., Fehr, D., & Fellner, G. (2004). Partner selection in public goods experiments. *Journal of Conflict Resolution*, *48*, 356-378. DOI: 10.1177/0022002704264143.

Duffy, J., & Feltovich, N. (1999). Does observation of others affect learning in strategic environments? An experimental study. *International Journal of Game Theory*, *28*, 131-152. DOI: 10.1007/s001820050102.

Duffy, J., & Ochs, J. (2009). Cooperative behavior and the frequency of social interaction. *Games and Economic Behavior*, *66*, 785-812. DOI: 10.1016/j.geb.2008.07.003.

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review*, *112*, 912-931. DOI: 10.1037/0033-295X.112.4.912.

Erev, I., Bereby-Meyer, Y., & Roth, A. E. (1999). The effect of adding a constant to all payoffs: experimental investigation and implications for reinforcement learning models. *Journal of Economic Behavior and Organization*, *39*, 111-128. DOI: 10.1016/S0167-2681(99)00028-1.

Erev, I., Ert, E., & Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making*, *21*, 575-597. DOI: 10.1002/bdm.602.

Feltovich, N., Iwasaki, A., & Oda, S. H. (2009). Payoff levels, loss avoidance, and equilibrium selection in the Stag Hunt: an experimental study. Working paper, University of Aberdeen.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*, 171-178. DOI: 10.1007/s10683-006-9159-4.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*, 397-404. DOI: 10.1016/S0165-1765(01)00394-9.

Fudenberg, D., & Levine, D. K. (1998). *The theory of learning in games (economic learning and social evolution)*. Cambridge, MA: MIT Press.

Grosskopf, B., Erev, I., & Yechiam, E. (2006). Foregone with the wind: indirect payoff information and its implications for choice. *International Journal of Game Theory*, *34*, 285-302. DOI: 10.1007/s00182-006-0015-8.

Huck, S., Müller, W., & Normann, H.-T. (2001). Stackelberg beats Cournot: on collusion and efficiency in experimental markets. *Economic Journal*, *111*, 749-765. DOI: 10.1111/1468-0297.00658.

Kagel, J. H., & Roth, A. E. (1995). *Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press.

Munichor, N., Erev, I., & Lotem, A. (2006). Risk attitude in small timesaving decisions. *Journal of Experimental Psychology: Applied*, *12*, 129-141. DOI: 10.1037/1076-898X.12.3.129.

Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, *8*, 164-212. DOI:10.1016/S0899-8256(05)80020-X.

Schmidt, D., Shupp, R., Walker, J. M., & Ostrom, E. (2003). Playing safe in coordination games: the roles of risk dominance, payoff dominance, and history of play. *Games and Economic Behavior*, *42*, 281-299. DOI:10.1016/S0899-8256(02)00552-3.

Siegel, S., & Castellan Jr., N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Van Huyck, J. B., Battalio, R. C., & Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, *80*, 234-248.

Appendix A: instructions from original experiment

Below is a annotated translation of the instructions given to subjects in this experiment. Copies of the actual instructions (in Japanese) are available from the corresponding author upon request.

1 Introduction

Thank you for your participation in an experiment in the economics of decision-making. If you follow these instructions carefully and make good decisions you might earn a considerable amount of money that will be paid to you in cash at the end of the session.

2 Sequence of Play in a Round

This experimental session consists of six different games. Each game consists of forty rounds. At the start of each game, you will be randomly assigned a player type, either “row player” or “column player.” Your type will not change during the course of game. In each game of this experiment you will be randomly matched to a player of the opposite type. You will be matched with a different player in every round [every fortieth round in the F treatment] of a game. We will refer to the person you are paired with in a round as your “partner.” Your score in each round will depend on your choice and the choice of your partner in that round. You will not know the identity of your partner in any round, even after the end of the session.

- At the beginning of each game, the computer program randomly matches each player to a partner.

- You and your partner play the game. Figure 1 is displayed on your screen. If you are a row player, you choose which row of the payoff table to play, R1 or R2. If you are a column player, you choose which column of the payoff table to play, C1 or C2.

- After all players have chosen actions, your action, your partner’s action, and your payoff or score are displayed. Your score is determined by your action and the action of your partner according to the given payoff table.

- *[This part is replaced according to the treatment.]*

  - **R treatment:** Provided that the last round of the game has not been reached, a new round of the same game will then begin. You will be matched with a different partner in the new round.
  - **F treatment:** Provided that the last round of the game has not been reached, a new round of the same game will then begin. You will be matched with the same partner in the new round. Nevertheless, at every forty rounds, your partner and the payoff table will be changed.

- **Notice that you must not record any results of the games.** If the experimenter find you are recording them, you cannot continue your experiment. In that case, you will not be paid for this experiment.

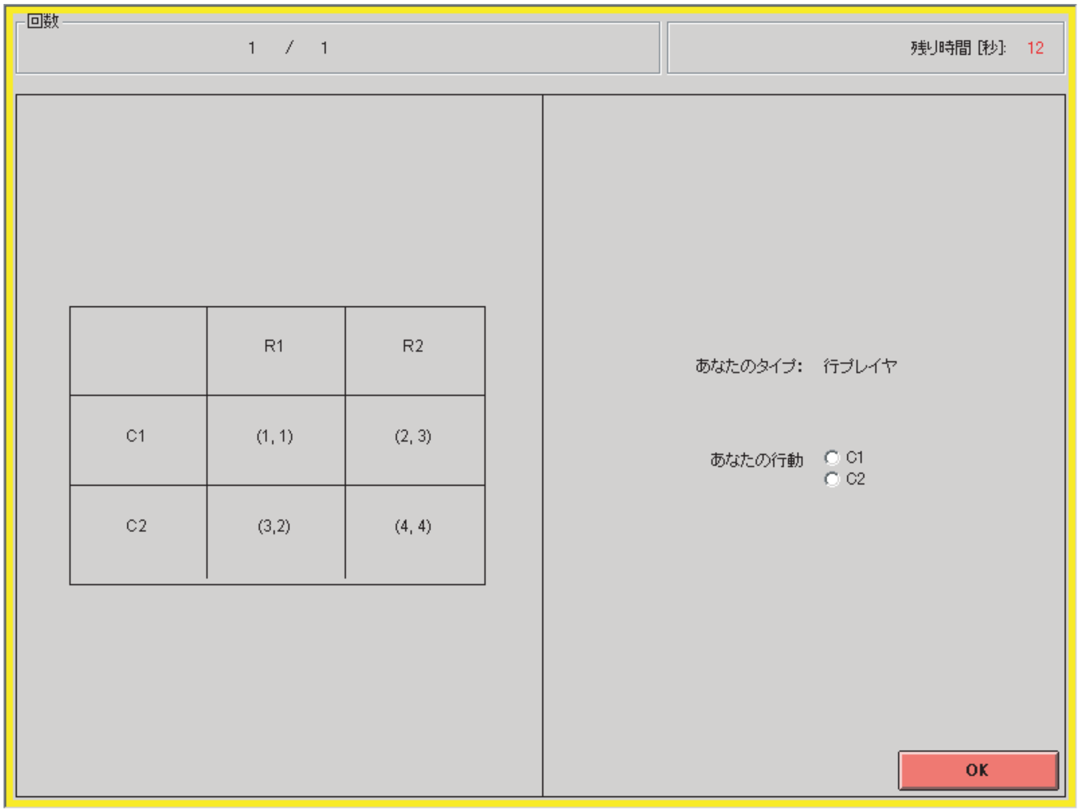


Figure 1: The screen when you are a column player. (The payoff table on the left side will not be shown)

3 The payoff tables

The payoff table for each game you play will not be shown on your computer screen. However, let us explain the payoff table to support your decision-making. In every round of a game, both you and your partner have a choice between two possible actions. If you are designated as the row player, you must choose between actions R1 and R2. If you are designated as the column player, you must choose between actions C1 and C2. Your action, together with the action chosen by your partner, determines one of the four boxes in the payoff table. In each box, the first number represents your score and the second number represents your partner's score.

4 Payments

If you complete this experiment, the computer screen will reveal your score, the round that you got the score, and the payment you obtain. The round will be randomly chosen from all rounds you played. The payment will be calculated from your score as 200 yen for each point in that round. In addition, you will be paid a show-up fee of 3000 yen.

Are there any questions before we begin?

Appendix B: Instructions from follow-up experiment

Below are the instructions used in the follow-up experiment. Text in square brackets appeared only in the F treatment, text in curly brackets appeared only in the R treatment, and text not in brackets appeared in both treatments.

You are about to participate in a decision-making experiment. Please read these instructions carefully, as the amount of money you earn may depend on how well you understand them. If you have a question at any time, please feel free to ask the experimenter. We ask that you not talk with the other participants during the experiment.

This experimental session consists of a game that is played for 50 rounds via your computer. In each round, you are matched to another participant, with whom you play the game. [The person matched to you will be chosen randomly by the computer program at the beginning of the experiment, and will remain the same for all rounds.] {The person matched to you will be chosen randomly by the computer program at the beginning of each round.} You will not be told the identity of the person matched to you, nor will he/she be told your identity – even after the end of the session.

The payoff table: In each round, both you and the person matched to you will have a choice between two possible actions, which will be called X and Y. Your action and the other person's action determine your earnings (in pence) for that round, according to a payoff table. The other person's earnings may be different from yours.

Example: This is an example of a payoff table. You will be using a different payoff table in the experiment, but it will have a similar structure. **The payoff table you actually use will be shown on your computer screen, but you will not see the numbers – they will be replaced by question marks.**

		Other person action	
		X	Y
Your action	X	Your earnings: 15	Your earnings: 25
	Y	Your earnings: 40	Your earnings: 0

Sequence of play in a round: The sequence of play in a round is as follows.

(1) [If it is the first round, the computer randomly matches you to another participant.] {The computer randomly matches you to another participant.}

(2) You and the person matched to you play the game. You choose an action, either X or Y. The person matched to you also chooses an action, either X or Y. Both of you make your choice without knowing the other's choice.

(3) The round ends. Your computer screen will display your choice, your earnings for the round, and your total earnings up to the current round. **You are not shown the choice or earnings of the person matched to you.**

Payments: At the end of the experimental session, you will be paid (in pence) the sum of your earnings in all rounds. Payments are made privately and in cash at the end of the session.