Learning Under Limited Information*

Yan Chen
School of Information, University of Michigan
550 East University Avenue, Ann Arbor, MI 48109-1092

Yuri Khoroshilov
Department of Finance, University of Michigan Business School
701 Tappan Street, Ann Arbor, MI 48109-1234

August 28, 2002

Running title: learning under limited information

Address for manuscript correspondence:
Yan Chen
School of Information, University of Michigan
550 East University Avenue, Ann Arbor, MI 48109-1092
Email: yanchen@umich.edu
Phone: (734) 763-2285
Fax: (734) 764-2475

*We thank Julie Cullen, Ido Erev, Dan Friedman, Eric Friedman, Drew Fudenburg, Amy Greenwald, Teck-Hua Ho, Lutz Kilian, David Lucking-Reiley, Atanasios Mitropoulos, Tom Palfrey, Rajiv Sarin, Scott Shenker, Farshid Vahid and seminar participants at Michigan, Mississippi, Rutgers, Texas A&M, the 1998 Econometric Society meetings (Chicago), ESA 2000 (New York City) and the 2001 Econometric Society meetings (New Orleans) for discussions and comments. We are grateful to John Van Huyck for providing the data and for helping us with part of the estimation. Emre Sucu provided excellent research assistance. We are grateful to two anonymous referees for their constructive comments which significantly improved the paper. The research support provided by NSF grant SBR-9805586 and SES-0079001 to Chen and the Deutsche Forschungsgemeinschaft through SFB303 at Universität Bonn are gratefully acknowledged. Any remaining errors are our own.

Abstract

We study how human subjects learn under extremely limited information. We use Chen's (forthcoming) data on cost sharing games, and Van Huyck, Battalio and Rankin's (1996) data on coordination games to compare three payoff-based learning models. Under the serial mechanism and coordination games, the payoff-assessment learning model (Sarin and Vahid, 1999) tracks the data the best, followed by the experience-weighted attraction learning model (Camerer and Ho, 1999), which in turn, is followed by a simple reinforcement learning model. Under the average cost pricing mechanism, however, none of the learning models tracks the data well.

Keywords: limited information, learning

1 Introduction

Limited information environments are found in many economic settings, for example, in distributed networks. In a distributed system such as the Internet agents have very limited *a priori* information about other agents and the payoff matrix (Friedman and Shenker, 1998). To design mechanisms for resource allocation in distributed systems it is important to study how agents learn in settings which have the characteristics of distributed systems. In this paper we study how human subjects learn under limited information in the laboratory.

From the perspective of studying human learning, limited information environments provide the simplest setting for evaluating learning in repeated games. When agents only know their own past actions and the corresponding own payoffs, belief learning models are not applicable. Learning models which depend on only a player's own actions and payoffs are relevant. These include simple reinforcement learning models (such as Roth and Erev, 1995), payoff-assessment learning models (Sarin and Vahid, 1999), a modified experience-weighted attraction learning model (Camerer and Ho, 1999) and responsive learning automata (Friedman and Shenker, 1995).

Several investigators have experimentally studied behavior in limited information environments. Mookherjee and Sopher (1994) found that in a repeated matching pennies game in which players did not know the other player's payoffs nor the other player's decisions, play was significantly different from the minimax solution. Van Huyck *et al.* (1996) report experimental results from a median-action coordination game under limited information and find that median player always converged to the interior Nash equilibrium and convergence was much faster than the Cross reinforcement learning model. Chen (forthcoming) studies the serial and average cost pricing mechanism under limited information and asynchronous actions. She finds that the serial mechanism significantly outperforms the average cost pricing mechanism under limited information even though both games are dominance-solvable. Friedman *et al.* (2000) report an experiment on learning in a continuous-time, limited information setting where subjects played the Cournot oligopoly game and the serial cost sharing game. Mitropoulos (forthcoming) reports an experiment on a 2×2 game of mutual fate control under limited information and finds that experimental results are unfavorable for the reinforcement learning rule.

In this paper we use Chen's experimental data on cost sharing games, and Van Huyck *et al.*'s data on coordination games to compare three payoff-based learning models. Section

2 introduces the two experiments used in the paper. Section 3 introduces three payoff-based learning models that we will evaluate. Section 4 reports parameter estimates, out-of-sample forecasting and the ranking of the three learning models based on the out-of-sample forecasting. Section 5 concludes with a discussion of various aspects of learning models which might predict well in limited information environments in light of our results.

2 The Experiments

We rank the learning models by using two samples of experimental data. These games are Chen's (forthcoming) cost sharing games, and Van Huyck *et al.*'s (1996) median-action coordination games. We chose these games for several reasons.

First, both experiments have the characteristics that the players have extremely limited information. Each player only knows his own feasible set of actions, his own past choices and the resulting own payoffs. As reviewed in the Introduction, there are not many experiments with this information condition.

Second, the games have a range of different structural features. In Chen (forthcoming) each of the two cost sharing games has a unique dominance solvable Nash equilibrium. However, the serial game is both dominance-solvable and overwhelmed-solvable, which is a stronger solution concept than dominance solvability (Friedman and Shenker, 1998), while the average cost pricing game is only dominance-solvable but not overwhelmed-solvable. In Van Huyck *et al.* (1996) each of the four coordination games have two strict Nash equilibria. This avoids the possible mistake of concluding that a model generally fits well because it happens to fit one class of games.

Third, the games have different spans - the cost sharing games last 150 rounds, while the coordination games last 40, 70 or 75 rounds. A mixture of long and short spans is a challenge to the learning models, which ought to be able to predict why convergence is quick in some games and slow in others.

Next we describe the main features and findings in each of the two experiments.

2.1 Cost Sharing Games in Chen (forthcoming)

Chen (forthcoming) studies the serial and average cost pricing mechanisms under limited information as well as complete information. We will only use the limited information data here.

Suppose a group of n agents share a one-input, one-output technology with decreasing returns. Each of the n agents announces his demand q_i of output. Let $q_1 \leq q_2 \leq \cdots \leq q_n$. The cost function is denoted by C, which is strictly convex. The cost sharing mechanism must allocate the total cost $C(\sum_i q_i)$ among the n agents. Under the serial mechanism (hereafter shortened as **SER**) agent 1 (with the lowest demand) pays $(1/n)$th of the cost of producing nq_1, $C(nq_1)/n$. Agent 2 pays agent 1's cost share plus $1/(n-1)$th of the incremental cost from nq_1 to $(n-1)q_2 + q_1$, i.e., $\frac{C(nq_1)}{n} + \frac{C(q_1+(n-1)q_2)-C(nq_1)}{n-1}$. And so on. Therefore, an agent's cost share is independent of demands higher than his own.

Under the average cost pricing mechanism (hereafter shortened as **ACP**), when agent i demands q_i amount of output, agent i's cost share is given by $(q_i/\sum_k q_k)C(\sum_k q_k)$, for all $i = 1, \cdots, n$. Therefore, an agent's cost share is proportional to his own demand. It is affected by his own demand and the sum of all other agents' demands.

In the experiment agents are endowed with linear preferences, $\pi_i(x_i, q) = \alpha_i q_i - x_i$, where α_i is agent i's marginal utility for the output, and x_i is his cost share. The cost function is quadratic, $C(q) = q^2$. There are two types of players, $\alpha_1 = 16.1$ and $\alpha_2 = 20.1$. The mechanisms are implemented as normal form games with a discrete strategy space, $S_i = \{1, 2, \cdots, 11, 12\}$ for each i. Under the serial mechanism the unique, dominance-solvable Nash as well as the Stackelberg equilibrium is $(4, 6)$. Under ACP, the unique, dominance-solvable Nash equilibrium is $(4, 8)$, while the Stackelberg equilibrium with player 2 as the leader is $(2, 12)$.

There are two treatments with synchronous play and updating, two with asynchronous play and updating. In the synchronous treatments (hereafter shortened as **SYN**) every player receives his own payoff feedback for each round. In the asynchronous treatments (hereafter shortened as **ASYN**) player 1 submits a demand and gets a payoff feedback each round, but player 2 submits a demand which is matched with his opponent's demands for the next five rounds and gets a cumulative payoff feedback every five rounds. Therefore, in the asynchronous treatment player 2 acts five times slower than player 1 and becomes the *de facto* Stackelberg leader. There are two types of matching protocols - random re-matching and fixed pairs. In all four treatments the game lasts for 150 rounds (30 rounds for player 2 in ASYN). We summarize the four treatments:

1. SYN_r: synchronous play, with random re-matching for each of the 150 rounds;

2. SYN_f: synchronous play, with repeated fixed pairs for 150 rounds;

3. ASYN_r: asynchronous play, with random re-matching for each of the 150 rounds; and

4. ASYN_f: asynchronous play, with repeated fixed pairs for 150 rounds.

The main findings are as follows. First, Kolmogorov-Smirnov test ($d = 0.093$ yielding a p-value of 73.6%, two-tailed) cannot reject the null hypothesis of uniform play in the initial period. This motivates our use of the uniform distribution as the initial distribution for the cost sharing games in Section 4. Second, under all four treatments the serial mechanism performs significantly better than the ACP mechanism both in terms of the proportion of equilibrium play and group efficiency. Third, the presence of asynchrony reduces the amount of equilibrium play.

2.2 Coordination Games in Van Huyck, Battalio and Rankin (1996)

Van Huyck *et al.* (1996) report an experiment on coordination games. Let e_i denote player i's action, and e_{-i} be the actions chosen by other players. The payoff function for player i is given by

$$\pi(e_i, e_{-i}) = 0.5 - |e_i - \omega M(e)(1 - M(e))|, \tag{1}$$

where $\omega \in (1, 4]$ and $M(e)$ is the median of e. The stage game has two strict Nash equilibria, $(e, M) = (0, 0)$ and $(1 - 1/\omega, 1 - 1/\omega)$, both of which are symmetric and efficient.

In the experiments, Van Huyck *et al.* (hereafter shortened as VHBR) implement four treatments by setting $\omega = 2.44, 2.47, 3.85$ and 3.87 respectively. Each individual i has the same set of 101 actions in each round, $E_i = \{0, 1, \cdots, 100\}$. Each cohort has five subjects playing the same coordination game for T periods, where $T = 40, 70$, or 75 in different treatments. In each period, each individual chooses one action and receives a payoff. In all treatments players have limited information and synchronous play.

The main findings are as follows. First, Kolmogorov-Smirnov tests fail to reject the null hypothesis of uniform play in the initial period. "The largest Kolmogorov T statistics is 0.18 and the critical value at the 5% level of statistical significance is 0.29. The Smirnov T statistics is 0.15 and the critical value at the 5 percent level of statistical significance is 0.8." (Van Huyck *et al.*, 1996, p. 8) Again, this motivates our use of the uniform distribution as the initial distribution for the coordination games in Section 4. Second, median player always converged to the interior Nash equilibrium. Third, convergence was extremely rapid, much faster than the Cross learning model.

Sarin and Vahid (1997) compare the payoff-assessment learning model, which will be introduced in Section 3, with the Cross learning model using the VHBR (1996) data. They find that in contrast to the Cross learning model the payoff-assessment model converges to the same equilibrium and in roughly the same number of rounds as the data. Note that the Cross learning model is a reinforcement type of learning model, in which the updating of choice probabilities is different from the simple reinforcement learning model characterized in Eq. (2) and (3) in Section 3.

3 Learning Models

The design of mechanisms for distributed systems relies on the actual dynamic learning behavior of humans in this context. Fudenburg and Levine (1998) provide an excellent survey for the vast theoretical and empirical literature on learning. Because of the extremely limited information, learning models which utilize only a player's own payoff information are relevant. We evaluate three such learning models: a simple reinforcement learning model, a modified experience-weighted attraction learning model, and a payoff-assessment learning model. In an earlier version we evaluated four learning models, these plus the responsive learning automata, which turned out to have the worst performance. Therefore, we dropped it in this version. These three models will be introduced in turn.

Let n be the number of actions for each player in each round. For simplicity we will omit all subscripts which represent player i. Therefore, $\pi_j(t)$ is the payoff of strategy j in round t; $p_j(t)$ is the probability that strategy j is chosen in round t; and r is the discount factor. Since the coordination game has a large strategy space, we incorporate similarity functions into all three learning models to represent the idea that an agent might use similarity among strategies to simplify the decision problem he faces. This idea was used by Rubinstein (1988), Gilboa and Schmeidler (1997), and more recently by Sarin and Vahid (1997) who build similarity functions into the payoff-assessment learning model. Following Sarin and Vahid (1997), we assume that strategies are naturally ordered by their labels and use the Bartlett similarity function, $f_{jk}(h,t)$, to denote the similarity between the played strategy k and an unplayed strategy j at period t:

$$f_{jk}(h,t) = \begin{cases} 1 - |j-k|/h, & \text{if } |j-k| < h, \\ 0, & \text{otherwise.} \end{cases}$$

The parameter h determines the $h-1$ unplayed strategies on either side of the played strategy to be updated. Note that when $h = 1$, $f_{jk}(1, t)$ degenerates into an indicator function which equals one if strategy j is chosen in round t and zero otherwise.

The **reinforcement learning (RL)** model is a model of "rote" learning, in which actions which do well in the past are more likely to be repeated in the future. Learning models in this spirit have a long history in biology and psychology. Their systematic application in experimental economics starts from Roth and Erev (1995). Erev and Roth (1998) show that the RL model tracks the data well across a wide variety of experimental games with unique, mixed strategy equilibria. Note that the amount of information in the two experiments is exactly the same as that required by the RL model, which bases its predictions solely on the individual payoff gains of that specific subject. Subjects do not form beliefs nor perform any optimization procedures. An often noted disadvantage of the RL model is its inability to lock on to the optimal strategy even with a long horizon.

Let $R_j(t)$ be the discounted payoff sum of an individual from choosing strategy j. It is sometimes called the propensity to choose strategy j. It begins with some prior value, $R_j(0)$. Updating is governed by

$$R_j(t) = rR_j(t-1) + f_{jk}(h, t)\pi_k(t). \tag{2}$$

The choice probability for an individual at round $t + 1$ is

$$p_j(t+1) = \frac{e^{\lambda R_j(t)}}{\sum_{l=1}^{n} e^{\lambda R_l(t)}}, \quad \forall j, \tag{3}$$

where $\lambda \geq 0$ helps to scale up or scale down the relative weights of the discounted payoff sums. The exponential functional form (logit) is chosen rather than the normal form (probit) or power form because of its ability to deal with negative payoffs. The working paper version of Chen and Tang (1998) evaluates all three functional forms of the RL model on public goods games and concludes that their performance is statistically indistinguishable. The exponential form has been used to study learning in games by Mookerjhee and Sopher (1994, 1997), Ho and Weigelt (1996), Fudenberg and Levine (1998), McKelvey and Palfrey (1995, 1998), and Camerer and Ho (1999).

The **experience-weighted attraction (EWA)** learning model (Camerer and Ho, 1999) incorporates the reinforcement approach with belief-based approaches. A key difference of EWA from RL is that EWA weighs hypothetical payoffs from unchosen strategies by a parameter, δ, and weighs the payoff actually received from the chosen strategy by an additional $1-\delta$. Therefore, unchosen strategies which would have yielded high payoffs are more likely to be chosen as well. In the experiments considered in this paper although the underlying payoff structure is unknown to the subjects they can still form beliefs about the possible payoffs. Let $\bar{\pi}_j(0)$ be an individual's initial belief about the possible payoff he might obtain from strategy j before playing the game. These beliefs can be updated as actual payoffs are received from different strategies. Updating is governed by the following rule,

$$\bar{\pi}_j(t) = \alpha f_{jk}(h,t)\pi_k(t) + (1-\alpha f_{jk}(h,t))\bar{\pi}_j(t-1),$$

where k is the strategy used in period t, and $\alpha = 0.75$ for the cost sharing games and $\alpha = 0.5$ for the coordination games. We tried other values of α and found these to be the best value among what we have searched.

The EWA model defines a variable, $N(t)$, which can be interpreted as the number of "observation-equivalents" of past experience, and a variable, $A_j(t)$, which measures the attraction of strategies. Both variables begin with some prior values, $N(0)$ and $A_j(0)$. Updating is governed by two rules. First,

$$N(t) = \rho N(t-1) + 1, \tag{4}$$

where ρ is a depreciation rate that measures the fractional impact of previous experience, compared to one new period. The second rule updates the level of attraction:

$$A_j(t) = \frac{N(t-1)A_j(t-1) + [\delta + (1-\delta f_{jk}(h,t))]\tilde{\pi}_j(t)}{N(t)}, \tag{5}$$

where $\tilde{\pi}_j(t) = f_{jk}(h,t)\pi_k(t) + (1-f_{jk}(h,t))\bar{\pi}_j(t)$.

The choice probability for an individual at round $t+1$ is

$$p_j(t+1) = \frac{e^{\lambda A_j(t)}}{\sum_{l=1}^{n} e^{\lambda A_l(t)}}, \quad \forall j. \tag{6}$$

The original EWA model has more parameters. Because of the limited information context and to make it comparable to other models, we reduce it to a three-parameter model. Because of this simplification, RL is no longer a special case of EWA.

The **payoff-assessment (PA)** learning model (Sarin and Vahid, 1999) is different from the above two learning models. It assumes that a player is a myopic subjective maximizer.

He chooses among alternate strategies only on the basis of the payoff he assesses he would obtain from them. These assessments do not explicitly take into account his subjective judgements regarding the likelihood of alternate states of the world. At each stage, the player chooses the strategy that he myopically assesses to give him the highest payoff and updates his assessment adaptively. Let $u_j(t)$ denote the subjective assessment of strategy s_j at time t. The initial assessments are denoted by $u_j(0)$. Payoff assessments are updated by taking a weighted average of his previous assessments and the objective payoff he actually obtains at time t. If strategy k is chosen at time t, then

$$u_j(t+1) = (1 - rf_{jk}(h,t))u_j(t) + rf_{jk}(h,t)\pi_k(t), \forall j. \tag{7}$$

Suppose that at time t the decision-maker experiences zero-mean, symmetrically distributed shocks, $Z_j(t)$ to his assessment of the payoff he would receive from choosing strategy s_j, for all s_j. Denote the vector of shocks by $Z = (Z_1, \cdots, Z_{12})$, and their realizations at time t by $z(t) = (z_1(t), \cdots, z_{12}(t))$. The decision maker makes choices on the basis of his shock-distorted subjective assessments, denoted by $\tilde{u}(t) = u(t)+Z(t)$. At time t he chooses strategy s_j if

$$\tilde{u}_j(t) - \tilde{u}_l(t) > 0, \forall s_l \neq s_j. \tag{8}$$

Note that mood shocks only affect his choices and not the manner in which assessments are updated. Sarin and Vahid (1999) prove that such a player converges to stochastically choose the strategy that first order stochastically dominates another among the strategies he converges to play with positive probability.

4 Parameter Estimation and Out-of-Sample Forecasting

We estimated the values of learning model parameters using two samples of experimental data and compared the learning models by their abilities to predict behavior out of sample.

For parameter estimation, we conducted Monte Carlo simulations designed to replicate the characteristics of each of the experimental settings. We then compare the simulated paths with the actual paths of a subset of the experimental data to estimate the parameters which minimize the mean-squared deviation (MSD) scores. Since the final outcome distributions of our data are unimodal, the simulated mean is an informative statistic and is well captured by MSD (Haruvy and Stahl, 2000).

In each simulation, 10,000 pairs of players were created, which yields a statistical accuracy of 1%. In each simulation the following steps were taken:

1. Initial values: Since Kolmogorov-Smirnov tests of the initial choice distribution by experimental subjects cannot reject the null hypothesis of uniform distribution in both data sets, in the cost sharing games we set $R_j(0) = A_j(0) = \bar{\pi}_j(0) = u_j(0) = 200$ for all players in RL, EWA and PA respectively, since the average first-round payoffs was around 200. This also results in a probability predictions around the centroid, $(1/12, \cdots, 1/12)$, for the first round. For the same reason, in the coordination games we set $R_j(0) = A_j(0) = \bar{\pi}_j(0) = u_j(0) = 0.5$ for all players in RL, EWA and PA respectively. The other variable of the EWA model, $N(0)$, was set to zero.

2. Simulated players were matched into fixed pairs, or randomly rematched pairs for each period, depending on the treatment.

3. The simulated players' strategies were randomly determined via Eq.(3) for RL, and Eq. (6) for EWA. Their strategies were determined via Eq. (8) for PA.

4. Payoffs were determined using the SER, ACP payoff rule or Eq. (1) employed in the experiment in question.

5. Propensities were updated according to Eq. (2) for RL. Number of observation-equivalents and attractions were updated according to Eq. (4) and Eq. (5) respectively. Assessments were updated according to Eq. (7) for PA. For cost-sharing games updating occurred every period under SYN for both players, every period for player 1 in ASYN and every five periods for player 2 in ASYN. For coordination games updating occurred every period.

The discount factor, $r \in [0, 1]$, and depreciation rate, $\rho \in [0, 1]$, were searched at a grid size of 0.1, and the scale factor, λ was searched at a grid size of 0.001 for the cost sharing games and 0.1 for the coordination games until the average MSD reached the minimum. Grid size was determined by the magnitude of payoffs.

Mood shocks in PA, z, were drawn from a uniform distribution on an interval $[-a, a]$. For the cost sharing games, a was searched on $[0, 100]$ with a step size of 10 to locate the best interval, and was then searched on this interval with a step size of 1 until the average MSD reached the minimum. For the coordination games, a was searched on $[0.02, 0.2]$ with a step size of 0.02 until the average MSD reached the minimum. Again the intervals and grid size were determined by the magnitude of payoffs.

To check the robustness of the PA model with respect to different distributions of mood shocks, we considered two other specifications and tested them on the cost sharing data: (1) z was drawn from an extreme value distribution with parameter a, which implies that the difference, $z_i(t) - z_j(t)$, has a logistic distribution. We call this model PA^l. (2) z was drawn from an exponential distribution with parameter a, which implies that the difference, $z_i(t) - z_j(t)$, has a double exponential distribution. We call this model PA^{2e}. In both cases, the parameter a was searched on $[1, 10]$ with a step size of one, since $a = 1$ was small enough given the magnitude of payoffs while $a = 10$ gives almost flat density functions. Note that with all three distributions (or any other distribution) the standard forms would be inappropriate because the magnitude of payoffs differ from experiment to experiment.

For the coordination games we searched the size of the similarity window at $h = 1, 6$ and 12 and found $h = 6$ fit the data the best. Since we use $h = 6$, the RL model allows some unchosen strategies to receive foregone payoffs. This is a generalization of the simple choice reinforcement learning model, in which unchosen strategies do *not* receive foregone payoffs.

We calibrated the models by deriving MSD estimates using nine independent pairs (three pairs from each subject pool) for each treatment in the cost sharing games, and two independent sessions for each treatment of the coordination games (cohorts 9 and 10 for $\omega = 3.87$, cohorts 15 and 16 for $\omega = 2.47$, cohorts 3 and 4 for $\omega = 3.85$ and cohorts 7 and 8 for $\omega = 2.44$). Then we validated the models by using the derived estimates to predict the path of play in the remaining independent sessions. This procedure uses enough data to estimate parameters reliably, and also have enough remaining data for out-of-sample forecasting. Camerer and Ho (1999) use the first 70% of the observations in each sample for parameter estimation and use the derived estimates to predict the path of play in the remaining 30% of the sample. Their approach is frequently used in time-series analysis in macroeconomics, because there is usually only one independent time-series observation. In contrast, in experimental economics we typically have several independent observations for each treatment. Since there is little reason to believe that learning patterns in earlier periods of the game remain the same as the later periods, our method seems to be more appropriate in the context of comparing learning models in experimental games.

As a baseline, the MSD for a **random choice model (RC)**, where each individual randomly chooses one of the alternatives with equal probability for all rounds, equals 0.916

for the cost sharing games and 0.990 for coordination games. Since this model is based on totally random choices of each subject, a learning model that produces a MSD greater than the above scores probably does not capture the essence of individual learning processes.

[Figures 1 and 2 about here.]

Figures 1 and 2 present the time series data from experimental subjects under SER and ACP respectively. Each graph presents the mean (the stars), standard deviation (the error bars) and stage game equilibria (the dashed lines) for each of the two different types averaged over all sessions for each mechanism. The four graphs in the first column display the average demand (and standard deviation) for type 1. The second column displays the average demand for type 2. These two figures reveal a fairly clear pattern of convergence under the serial mechanism, but not under ACP.

[Table I about here.]

Table I reports the calibration results in the SER games. We took eighteen fixed pairs in the SER games and calculated the estimated parameter values and MSD scores. In the PA model mood shocks are drawn from a uniform interval of [-3, 3] under SYN and [-2,2] under ASYN, which is very small compared with the magnitude of payoffs. Compared with SYN, the introduction of asynchrony increases the discount factor, r, to one. Intuitively, asynchrony as implemented in Chen (forthcoming) stabilizes the play by allowing player 2 to change his message only once every five rounds, therefore we have smaller shocks and larger discount factors. For the EWA model, in both SYN and ASYN, we have $\delta = 1$, which indicates that player's belief about payoffs associated with a strategy completely outweighs the actual received payoffs. This could also be due to the fact that play converged rather quickly in SER. For the same reason, the depreciation parameter, $\rho = 0.1$, is rather small. In RL the discount factor equals 1.0 in both cases. This indicates that RL keeps track of all past payoffs. With discount factor less than one, convergence of RL becomes much slower. In both EWA and RL increasing λ increases fit. In each case we report the largest λ that does not cause explosion.

[Table II about here.]

Table II reports the validation results, using the best fit parameters in Table I on the hold-out sample, including the fixed pair treatment and the random matching treatment.

[Tables III and IV about here.]

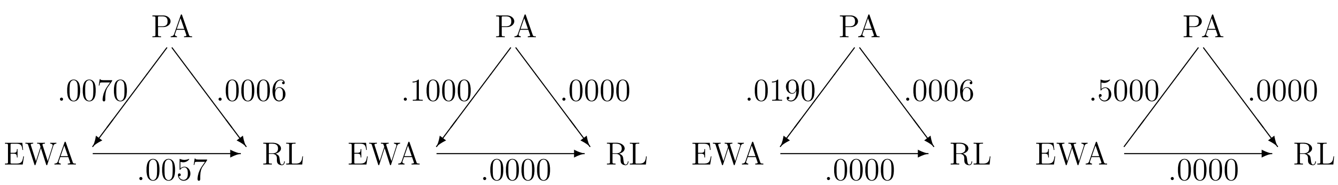
Table III reports the calibration results in SER and ACP games using two other versions of PA: PA^l and PA^{2e}. The estimated parameters have similar patterns as those of the original PA model: (1) from SYN to ASYN, the discount factor, r, increases; (2) Parameter, a, is much larger under ACP than under SER, indicating much larger shocks under ACP because of its volatile paths. Table IV reports the validation results, using the best fit parameters in Table III on the hold-out sample. Compared to the corresponding columns 2 and 3 in Tables II and VI, we can see that the MSD scores are very similar under the three versions of the PA model. Indeed, this will be confirmed in Results 1 and 2 using permutation tests.

In presenting the results, we introduce two shorthand notations. First, $x \rightarrow y$, denotes x "beats" y, i.e., MSD(x) is significantly less than MSD(y) at the 10% level or less. Second, $x—y$ denotes that the difference between MSD(x) and MSD(y) is statistically insignificant at the 10% level.

RESULT 1 (Learning Models under SER) : *The ranking of MSD of the three learning models is highly significant under the serial mechanism: PA → EWA → RL.*
The ranking of MSD of the three versions of the PA model is statistically insignificant under the serial mechanism except one case: under $ASYN_r$, $PA^l \rightarrow PA$ at the 10% level.

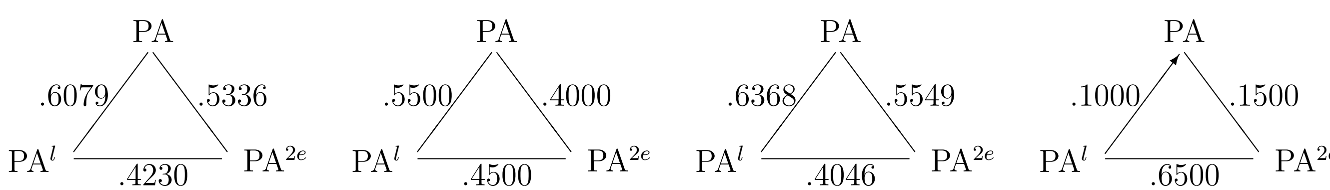
SUPPORT. P-values from one-tailed permutation tests using the MSD scores in Table II are summarized in the following four triangles, one for each treatment:

SER-SYN_f: SER-SYN_r: SER-$ASYN_f$: SER-$ASYN_r$:



P-values from one-tailed permutation tests using the MSD scores in Table IV are summarized in the following four triangles, one for each treatment:

SER-SYN_f: SER-SYN_r: SER-$ASYN_f$: SER-$ASYN_r$:



Therefore, under the serial mechanism PA clearly outperforms every other model, followed by EWA, which in turn, is followed by RL. It is interesting to note that the degree of optimization in each learning model follows virtually the same order. Even in cases with extremely limited information experimental subjects seemed to have carried out a fair amount of optimization, as embodied by PA and EWA.

[Figure 3 about here.]

Figure 3 presents the simulated time series paths under SERSYN_f from the PA, EWA and RL models respectively. Simulated paths under other treatments of the SER games exhibit similar patterns. Comparing the Monte Carlo simulations with the corresponding real data in the first row of Figure 1 one can see that under the serial mechanism experimental subjects converged a lot faster to the equilibrium prediction than that predicted by the RL model. The PA and EWA models capture the convergence dynamics much better.

In the ACP games, the subjects did not converge to the equilibrium strategies. As analyzed in Section 2, in the SER games a player's payoffs are independent of demands larger than his own, while under ACP a player's payoffs are affected by every player's demands. The greater the variation a player can have on the payoffs of other players, the less inherently stable the play of the game is. Under ACP one player's experimentation is more likely to cause experimentation of other players. Therefore, the payoff feedbacks in the ACP game are not as useful as the SER game. This type of feedback which lead to more confusion and lack of stability is called "experimentation cascades" by Friedman *et al.* (2000). As revealed by Figure 2, the ACP data exhibited relatively volatile paths. This turns out to be a challenge for the learning models.

[Table V about here.]

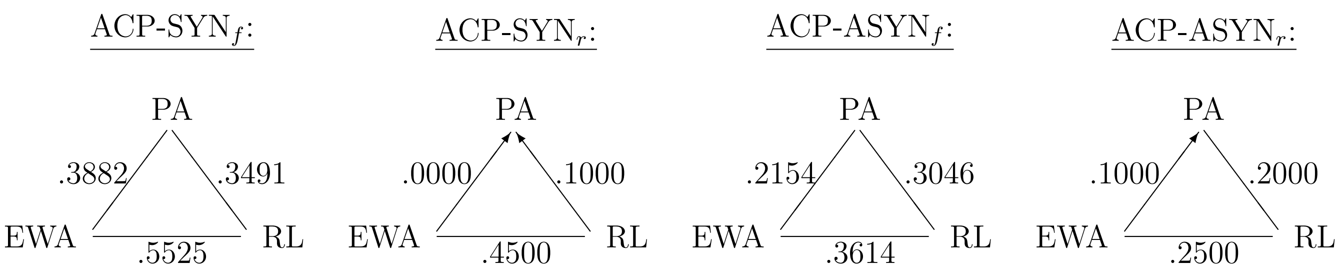
Table V reports the calibration results in the ACP games. We took eighteen fixed pairs in the ACP games and estimated the parameter values and MSD scores. In the PA model mood shocks are drawn from a uniform interval of [-50, 50] under SYN and [-45, 45] under ASYN, which are much larger than that in the SER games, reflecting the volatile dynamic paths in the data. Compared with SYN, the introduction of asynchrony again increases the discount factor, r, from 0.4 to 0.8. This is mirrored in EWA by an increase of the depreciation parameter, ρ, from 0.1 to 0.4, and an increase of δ from 0.2 to 1.0. The discount factor in RL remains stable at 1.0.

[Table VI about here.]

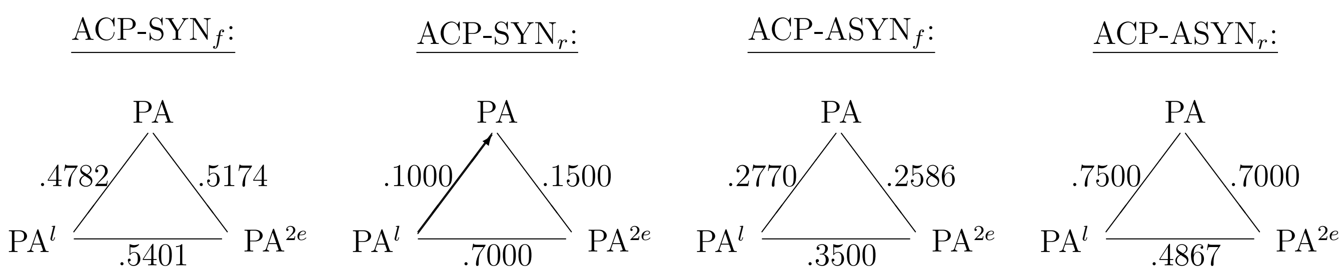
Table VI reports the validation results, using the best fit parameters in Table V on the hold-out sample, including the fixed pair treatment and the random matching treatment.

RESULT 2 (Learning Models under ACP) : *Under ACP the MSD of the three learning models are all close to the random choice prediction. The ranking of MSD of the three models is statistically insignificant except for three cases: under SYN_r: EWA $\rightarrow$ PA, and RL $\rightarrow$ PA; under $ASYN_r$: EWA $\rightarrow$ PA at the 10% level.*
The ranking of MSD of the three versions of the PA model is statistically insignificant under the ACP except one case: under SYN_r, $PA^l \rightarrow$ PA at the 10% level.

SUPPORT. P-values from one-tailed permutation tests using the MSD scores in Table VI are summarized in the following four triangles, one for each treatment:



P-values from one-tailed permutation tests using the MSD scores in Table IV are summarized in the following four triangles, one for each treatment:



Note that even though EWA beats PA in ACP games with random matching, the differences in the MSD scores are very small. Since there are only three independent observations in each random matching treatment while there are many more independent observations in the fixed-pairs treatments, we think that the results for the fixed-pair treatments are much more robust, i.e., the performance of the three learning models under ACP are statistically indistinguishable, and close to the prediction of the random choice model.

[Figure 4 about here.]

Figure 4 presents the Monte Carlo simulation results for the three learning models under ACP. Comparing the simulated paths with the corresponding data in the first row of Figure 2, one can see that although all three models capture the non-convergence patterns in the experimental data none of the models does a good job of describing how individual players learn and adapt.

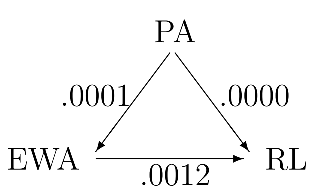
The ACP mechanism provides a challenging data set for the learning models. So far none of the three models does a good job of tracking the data. As discussed earlier ACP does not facilitate learning under limited information because each player's payoffs are affected by everyone else's demands. One player's experimentation immediately changes everyone else's payoffs, and thus can cause much more experimentation, as indicated by the data. In order to predict how subjects learn in ACP games, a learning model's noise (or error term) should depend on the previous strategies and corresponding payoffs in a hill-climbing fashion rather than being random. Intuitively, the direction and magnitude of experimentation ought to depend on the payoffs a player received from each strategy. This aspect is not captured by any of the existing learning models that we are aware of.

In the coordination games, we did two sets of analysis. In this game each player has 101 actions in each round. We use similarity windows to see whether it improves the predictions of various learning models.

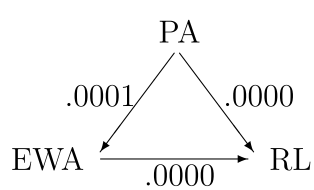
RESULT 3 (Learning Models under Coordination Games) : *The ranking of MSD of the three learning models is highly significant under the coordination games: PA → EWA → RL.*

SUPPORT. Results from one-tailed permutation tests using the MSD scores in Table VIII and X are summarized in the following two triangles:

VHBR: without similarity windows



VHBR: with similarity windows



Result 3 confirms Result 1 in the ranking of learning models.

RESULT 4 (Effect of Similarity Functions) : *The use of similarity function significantly improves the predictions of all three learning models in coordination games.*

SUPPORT. Comparison of validation results from Table VIII and X show that for each independent observation and each learning model, MSD with similarity windows is smaller than that without similarity windows. Any statistical test is superfluous.

Result 4 shows that subjects use the order of an unchosen strategy with respect to the chosen strategy to estimate its foregone payoffs and then use it to change their choices. This implies that learning models that do not use foregone payoffs might not work so well even in environments with limited information.

[Figures 5 and 6 about here.]

Figures 5 and 6 present the simulated paths for all treatments of the coordination games under each of the three learning models without and with similarity windows. A comparison of these two figures illustrates that adding similarity windows increases the speed of convergence under all three learning algorithms, and thus capturing the learning dynamics in the data a lot better.

5 Conclusion

To understand individual behavior under limited information three payoff-based learning models are evaluated by comparing Monte Carlo simulations of the learning models with experimental data. Under the serial mechanism and the coordination games the payoff-assessment learning model tracks the data the best, followed by the experience-weighted attraction learning model, which in turn, is followed by a simple reinforcement learning model. Learning models which incorporate more optimization, such as the payoff-assessment learning model and the experience-weighted attraction learning model, make better predictions under the serial mechanism and the coordination games. Under the average cost pricing mechanism, however, none of the learning models does a good job of tracking the data. To test robustness of the performance of the payoff-assessment learning model, we used three different specifications of the model and found their performance to be statistically indistinguishable.

Understanding individual learning behavior under limited information is important for mechanism design for distributed systems such as the Internet. The SER and ACP mechanisms are both used for congestion allocation on Internet routers. In the context of several agents sharing a network link, the cost to be shared is congestion experienced. Each agent controls the rate at which he is transmitting data, which corresponds to the demand in the cost sharing framework. If the sum of the transmission rates is greater than the total link capacity, then the link becomes congested and the agents' packets experience delays. Most current Internet routers use a FIFO packet scheduling algorithm, which results in each agent's average queue being proportional to his transmission rate. His queue is affected by everyone else's transmission rates. This corresponds to the average cost pricing mechanism (Shenker, 1990). In contrast, the Fair Queueing packet scheduling algorithm, which corresponds to the serial mechanism, leads to congestion allocations such that an agent's average queue is independent of transmission rates higher than his own. The new generation of Cisco 7200, 3600 and 2600 routers have both the FIFO and Fair Queueing options. Under the Fair Queueing algorithm, the environment is much more stable, and thus facilitating learning. Under FIFO, each agent's experimentation immediately affects everyone else's queue, which prompts more experimentation, therefore, the environment is not so stable.

A PA type of learner in the Internet setting would choose among transmission rates (or strategies) with the shortest delays (or best payoffs). Bad strategies are quickly abandoned. Therefore, in a stable environment under the Fair Queueing algorithm the agent with the smallest demand can quickly find the best strategy via hill-climbing and settle on that strategy or a close neighborhood of that strategy, the agent with the second smallest demand then finds the best strategy and settles, and so on. The whole system will converge rather quickly. In contrast, an EWA or RL learner will keep all strategies alive, therefore convergence is not as rapid. The original EWA model is designed to deal with situations where there is a role for beliefs. This is probably why the modified version does not do as well in limited information settings. The RL learner has too little optimization built in and therefore has too much inertia. In an unstable environment under FIFO (or ACP), the feedback is not as useful as in the SER game since it is the aggregate result of all agents' experimentation. A PA learner will be responding to all other learners' experimentation in trying to locate the best strategy. There is no obvious order of settling as under the SER mechanism. To capture how real humans learn in such an unstable environment, we speculate that the direction and magnitude of a learner's experimentation (or the noise term) should be a function of the payoffs received from each strategy. However, this aspect is not captured by any of the learning models we are aware of.

References

Camerer, C. and Ho, T. (1999). "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica* **67**, 827-874.

Chen, Y. (forthcoming). "An Experimental Study of Serial and Average Cost Pricing Mechanisms," *Journal of Public Economics*.

Chen, Y. and Tang, F. (1998). "Learning and Incentive Compatible Mechanisms for Public Goods Provision: An Experimental Study," *Journal of Political Economy* **106**, 633-662.

Erev, I. and Roth, A. (1998). "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review* **88**, 848-881.

Friedman, E., Shor, M., Shenker, S. and Sopher, B. (2000). "An Experiment on Learning with Limited Information: Nonconvergence, Experimentation Cascades, and the Advantage of Being Slow ," manuscript. New Brunswick: Rutgers University, Department of Economics.

Friedman, E. and Shenker, S. (1995). "Synchronous and Asynchronous Learning by Responsive Learning Automata," manuscript. New Brunswick: Rutgers University, Department of Economics.

Friedman, E. and Shenker, S. (1998). "Learning and Implementation on the Internet," manuscript. New Brunswick: Rutgers University, Department of Economics.

Fudenberg, D. and Levine, D. (1998). *Theory of Learning in Games*. Cambridge, MA: MIT Press.

Gilboa, I., and Schmeidler, D. (1997). "Act Similarity in Case-Based Decision Theory," *Economic Theory* **9**, 47-61.

Haruvy, E. and Stahl, D. (2000). "Initial Conditions, Adaptive Dynamics, and Final Outcomes: An Approach to Equilibrium Selection," manuscript. The University of Texas at Austin.

Ho, T. and Weigelt, K. (1996). "Task Complexity, Equilibrium Selection, and Learning: An Experimental Study," *Management Science* **42**, 659-679.

Mitropoulos, A. (forthcoming). "Learning under Minimal Information: An Experiment on Mutual Fate Control," *Journal of Economic Psychology.*

McKelvey, R. and Palfrey T. (1995). "Quantal Response Equilibria for Normal Form Games," *Games Econ. Behavior* **10**, 6-38, doi:10.1006/game.1995.1023.

McKelvey, R. and Palfrey, T. (1998). "Quantal Response Equilibria for Extensive Form Games," *Experimental Economics* **1**, 9-41.

Mookherjee, D. and Sopher, B. (1994). "Learning Behavior in an Experimental Matching Pennies Game," *Games Econ. Behavior* **7** : 62-91, doi:10.1006/game.1994.1037.

Mookherjee, D. and Sopher, B. (1997). "Learning and Decision Costs in Experimental Constant-Sum Games," *Games Econ. Behavior* **19**, 97-132, doi:10.1006/game.1997.0540.

Roth, A. and Erev, I. (1995). "Learning in Extensive Games: Experimental Data and Simple Dynamic Model in the Intermediate Term," *Games Econ. Behavior* **8**, 164-212.

Rubinstein, A. (1988). "Similarity and Decision-Making under Risk (Is There a Utility Theory Resolution to the Allais Paradox)," *Journal of Economic Theory* **46**, 145-153.

Sarin, R. and Vahid, F. (1999). "Payoff Assessments without Probabilities: A Simple Dynamic Model of Choice," *Games Econ. Behavior* **28**, 294-309, doi:10.1006/game.1998.0702.

Sarin, R. and Vahid, F. (1997). "Payoff Assessments without Probabilities: Incorporating 'Similarity' among Strategies," manuscript. College Station: Texas A&M University, Department of Economics.

Shenker, S. (1990). "Making Greed Work in Networks: A Game-Theoretic Analysis of Gateway Service Disciplines," manuscript. Palo Alto: Xerox Research Center.

Van Huyck, J., Battalio, R. and Rankin, F. (1996). "Selection Dynamics and Adaptive Behavior without Much Information," manuscript. College Station: Texas A&M University, Department of Economics.

Table I: Calibration in the SER Game: MSD and Parameter Values

	PA		EWA		RL	
Games	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$
MSD	0.346	0.409	0.853	0.617	0.748	0.754
	0.873	0.716	0.655	0.659	0.905	0.775
	0.338	0.539	0.790	0.660	0.749	0.781
	0.325	0.348	0.639	0.626	0.747	0.754
	0.390	0.675	0.634	0.722	0.757	0.791
	1.139	0.272	0.847	0.625	0.833	0.752
	0.444	0.455	0.639	0.647	0.757	0.768
	0.428	0.369	0.649	0.613	0.766	0.753
	0.594	0.675	0.665	0.665	0.772	0.784
Estimated Parameter Values	$r = 0.80$ $a = 3.00$	$r = 1.00$ $a = 2.00$	$\delta = 1.00$ $\rho = 0.10$ $\lambda = 0.007$	$\delta = 1.00$ $\rho = 0.10$ $\lambda = 0.007$	$r = 1.00$ $\lambda = 0.008$	$r = 1.00$ $\lambda = 0.008$

Table II: Validation in the SER Game

	PA		EWA		RL	
# of Sessions	SYN_r	$ASYN_r$	SYN_r	$ASYN_r$	SYN_r	$ASYN_r$
1	0.650	0.743	0.724	0.726	0.805	0.809
2	0.511	0.703	0.683	0.747	0.779	0.817
3	0.702	0.733	0.724	0.709	0.801	0.802
Overall	0.621	0.726	0.710	0.727	0.795	0.804
# of Pairs	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$
1	0.390	0.326	0.845	0.605	0.756	0.744
2	0.325	0.388	0.640	0.599	0.749	0.746
3	0.428	0.642	0.647	0.682	0.752	0.791
4	0.329	0.613	0.666	0.620	0.750	0.757
5	0.561	0.822	0.675	0.732	0.784	0.794
6	0.463	0.353	0.653	0.642	0.769	0.760
7	0.813	0.376	0.784	0.641	0.827	0.754
8	1.008	0.485	0.842	0.646	0.872	0.767
9	0.522		0.674		0.776	
10	0.399		0.641		0.761	
11	0.392		0.633		0.758	
Overall	0.512	0.501	0.700	0.646	0.778	0.764

Table III: Calibration in the SER and ACP Game: PA^l and PA^{2e}

	SER: PA^l		ACP: PA^l		SER: PA^{2e}		ACP: PA^{2e}	
Games	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$
MSD	0.393	0.458	0.898	0.844	0.365	0.435	0.899	0.848
	0.851	0.683	0.871	0.824	0.851	0.690	0.874	0.828
	0.387	0.565	0.851	0.930	0.359	0.550	0.859	0.923
	0.376	0.425	0.898	0.861	0.347	0.392	0.898	0.860
	0.433	0.646	0.886	0.972	0.408	0.648	0.885	0.965
	1.036	0.367	0.907	0.855	1.087	0.328	0.906	0.857
	0.472	0.492	0.897	0.847	0.456	0.472	0.898	0.848
	0.465	0.436	0.916	0.868	0.441	0.406	0.914	0.868
	0.585	0.662	0.939	0.827	0.591	0.662	0.936	0.830
Parameter Values	$r = 0.80$, $a = 1$	$r = 1.00$, $a = 1$	$r = 0.50$, $a = 10$	$r = 0.90$, $a = 6$	$r = 0.80$, $a = 1$	$r = 1.00$, $a = 1$	$r = 0.40$, $a = 10$	$r = 0.90$, $a = 8$

Table IV: Validation in the SER and ACP Game: PA^l and PA^{2e}

	SER: PA^l		ACP: PA^l		SER: PA^{2e}		ACP: PA^{2e}	
# of Sessions	SYN_r	$ASYN_r$	SYN_r	$ASYN_r$	SYN_r	$ASYN_r$	SYN_r	$ASYN_r$
1	0.648	0.707	0.896	0.888	0.645	0.713	0.896	0.888
2	0.528	0.684	0.893	0.892	0.513	0.684	0.893	0.891
3	0.689	0.701	0.879	0.910	0.694	0.706	0.881	0.908
Overall	0.622	0.697	0.889	0.897	0.618	0.701	0.890	0.896
# of Pairs	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$
1	0.434	0.396	0.844	0.851	0.411	0.365	0.852	0.852
2	0.377	0.452	0.836	0.820	0.349	0.423	0.843	0.825
3	0.461	0.648	0.839	0.873	0.435	0.640	0.847	0.871
4	0.381	0.597	0.885	0.945	0.352	0.597	0.887	0.939
5	0.579	0.754	0.899	0.846	0.567	0.771	0.900	0.848
6	0.497	0.424	0.897	0.836	0.474	0.392	0.896	0.838
7	0.784	0.434	0.926	0.945	0.788	0.406	0.922	0.941
8	0.954	0.516	0.919	0.886	0.978	0.498	0.916	0.883
9	0.541		0.860	0.863	0.529		0.865	0.862
10	0.438		0.943		0.416		0.935	
11	0.434				0.409			
Overall	0.535	0.528	0.885	0.874	0.519	0.512	0.886	0.873

Table V: Calibration in the ACP Game: MSD and Parameter Values

	PA		EWA		RL	
Games	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$
MSD	0.900	0.862	0.888	0.843	0.892	0.853
	0.870	0.821	0.857	0.823	0.845	0.831
	0.851	0.968	0.845	0.950	0.842	0.950
	0.898	0.858	0.895	0.851	0.894	0.856
	0.885	0.949	0.880	0.939	0.880	0.898
	0.910	0.874	0.903	0.868	0.905	0.870
	0.899	0.875	0.898	0.865	0.901	0.875
	0.917	0.868	0.921	0.864	0.928	0.869
	0.945	0.857	0.937	0.841	0.943	0.860
Estimated Parameter Values	$r = 0.40$ $a = 50$	$r = 0.80$ $a = 45$	$\delta = 0.20$ $\rho = 0.10$ $\lambda = 0.005$	$\delta = 1.00$ $\rho = 0.40$ $\lambda = 0.005$	$r = 1.00$ $\lambda = 0.008$	$r = 1.00$ $\lambda = 0.008$

Table VI: Validation in the ACP Game

	PA		EWA		RL	
# of Sessions	SYN_r	$ASYN_r$	SYN_r	$ASYN_r$	SYN_r	$ASYN_r$
1	0.901	0.887	0.892	0.862	0.894	0.879
2	0.900	0.889	0.889	0.863	0.889	0.866
3	0.893	0.908	0.871	0.904	0.871	0.904
Overall	0.898	0.895	0.884	0.876	0.885	0.883
# of Pairs	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$	SYN_f	$ASYN_f$
1	0.843	0.860	0.834	0.848	0.819	0.855
2	0.834	0.846	0.823	0.822	0.807	0.836
3	0.837	0.900	0.828	0.877	0.813	0.879
4	0.886	0.942	0.878	0.930	0.872	0.922
5	0.904	0.843	0.903	0.839	0.912	0.852
6	0.898	0.887	0.895	0.861	0.898	0.878
7	0.929	0.955	0.926	0.940	0.932	0.936
8	0.922	0.882	0.916	0.869	0.921	0.884
9	0.859	0.860	0.855	0.851	0.848	0.851
10	0.945		0.947		0.953	
Overall	0.886	0.886	0.880	0.871	0.877	0.877

Table VII: Calibration (Without Similarity Windows) in Coordination Games

Features		PA		EWA		RL	
Treatment	Cohort	MSD	Parameter	MSD	Parameter	MSD	Parameter
					$\delta = 1.0$		
$\omega = 2.44$	5	0.935	$r = 0.20$	0.951	$\rho = 0.3$	0.983	$r = 1.0$
(75 rounds)	6	0.929	$a = 0.16$	0.948	$\lambda = 1.5$	0.983	$\lambda = 1.0$
					$\delta = 1.0$		
$\omega = 2.47$	15	0.952	$r = 0.50$	0.980	$\rho = 0.1$	0.989	$r = 1.0$
(40 rounds)	16	0.956	$a = 0.14$	0.981	$\lambda = 1.5$	0.989	$\lambda = 1.0$
					$\delta = 1.0$		
$\omega = 3.85$	1	0.912	$r = 0.80$	0.953	$\rho = 0.1$	0.985	$r = 1.0$
(75 rounds)	2	0.938	$a = 0.16$	0.956	$\lambda = 1.5$	0.984	$\lambda = 1.0$
					$\delta = 1.0$		
$\omega = 3.87$	9	0.927	$r = 0.90$	0.983	$\rho = 0.1$	0.989	$r = 1.0$
(40 rounds)	10	0.935	$a = 0.14$	0.983	$\lambda = 1.5$	0.989	$\lambda = 1.0$

Table VIII: Validation (Without Similarity Windows) in Coordination Games

Treatment	Cohort	PA	EWA	RL
$\omega = 2.44$	7	0.930	0.946	0.983
(75 rounds)	8	0.931	0.945	0.983
$\omega = 2.47$				
(40 rounds)	-	-	-	-
$\omega = 3.85$	3	0.943	0.961	0.985
(75 rounds)	4	0.943	0.959	0.984
	11	0.931	0.983	0.989
	12	0.932	0.983	0.989
$\omega = 3.87$	13	0.927	0.983	0.989
(40 rounds)	14	0.927	0.983	0.988

Table IX: Calibration (With Similarity Windows) in Coordination Games

Features		PA		EWA		RL	
Treatment	Cohort	MSD	Parameter	MSD	Parameter	MSD	Parameter
					$\delta = 0.9$		
$\omega = 2.44$	5	0.916	$r = 0.9$	0.939	$\rho = 0.1$	0.975	$r = 1.0$
(75 rounds)	6	0.909	$a = 0.1$	0.933	$\lambda = 1.5$	0.975	$\lambda = 1.0$
					$\delta = 1.0$		
$\omega = 2.47$	15	0.950	$r = 1.0$	0.968	$\rho = 0.1$	0.982	$r = 1.0$
(40 rounds)	16	0.955	$a = 0.2$	0.971	$\lambda = 1.5$	0.983	$\lambda = 1.0$
					$\delta = 0.9$		
$\omega = 3.85$	1	0.872	$r = 0.8$	0.929	$\rho = 0.1$	0.972	$r = 1.0$
(75 rounds)	2	0.924	$a = 0.02$	0.939	$\lambda = 1.5$	0.974	$\lambda = 1.0$
					$\delta = 1.0$		
$\omega = 3.87$	9	0.912	$r = 1.0$	0.962	$\rho = 0.1$	0.979	$r = 1.0$
(40 rounds)	10	0.922	$a = 0.06$	0.962	$\lambda = 1.5$	0.980	$\lambda = 1.0$

Table X: Validation (With Similarity Windows) in Coordination Games

Treatment	Cohort	PA	EWA	RL
$\omega = 2.44$	7	0.906	0.934	0.975
(75 rounds)	8	0.906	0.934	0.975
$\omega = 2.47$				
(40 rounds)	-	-	-	-
$\omega = 3.85$	3	0.921	0.945	0.976
(75 rounds)	4	0.924	0.944	0.974
	11	0.917	0.962	0.979
	12	0.919	0.962	0.980
$\omega = 3.87$	13	0.913	0.962	0.979
(40 rounds)	14	0.912	0.962	0.979

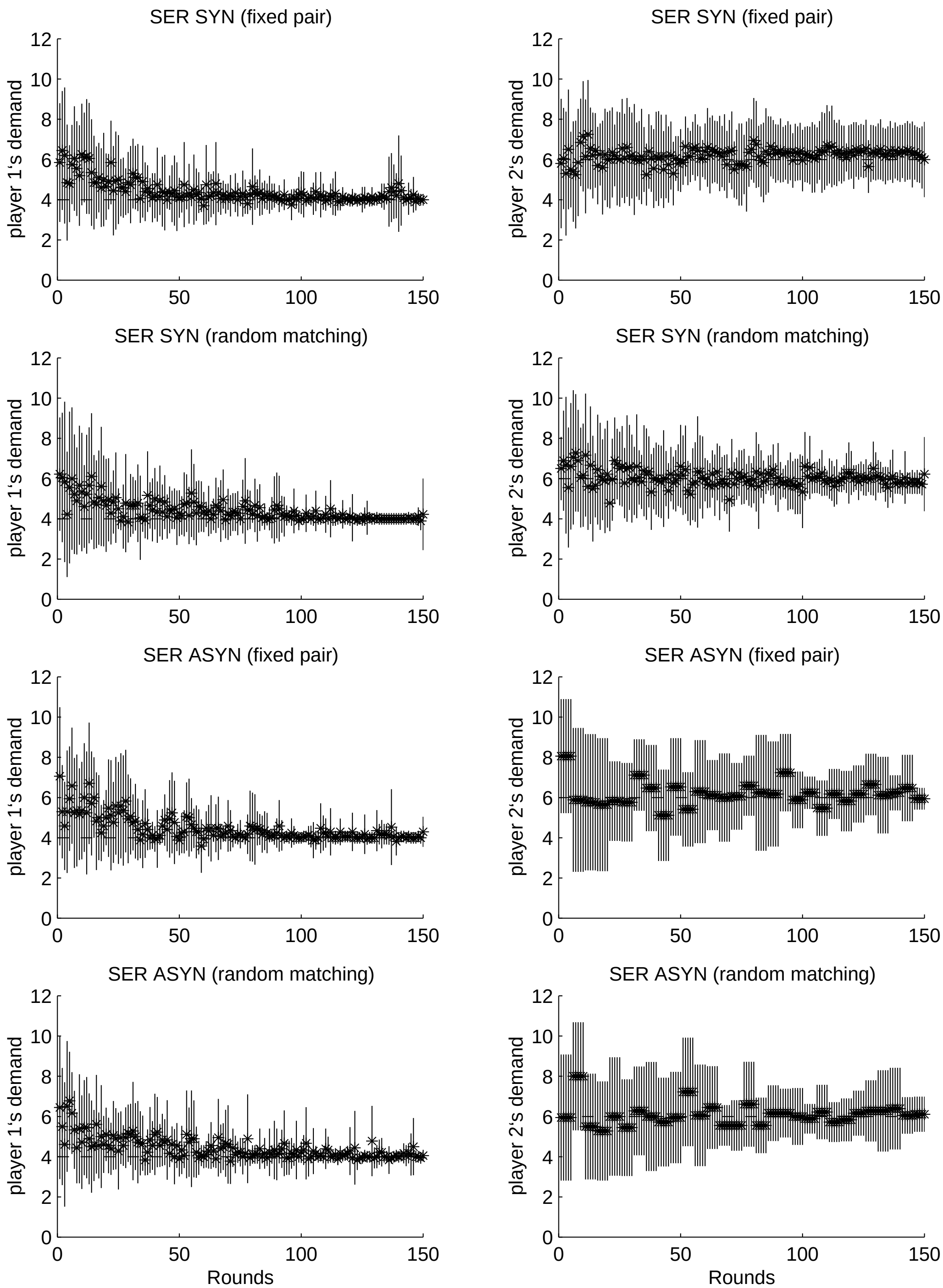


Figure 1. Experimental Data under SER in Chen (forthcoming).

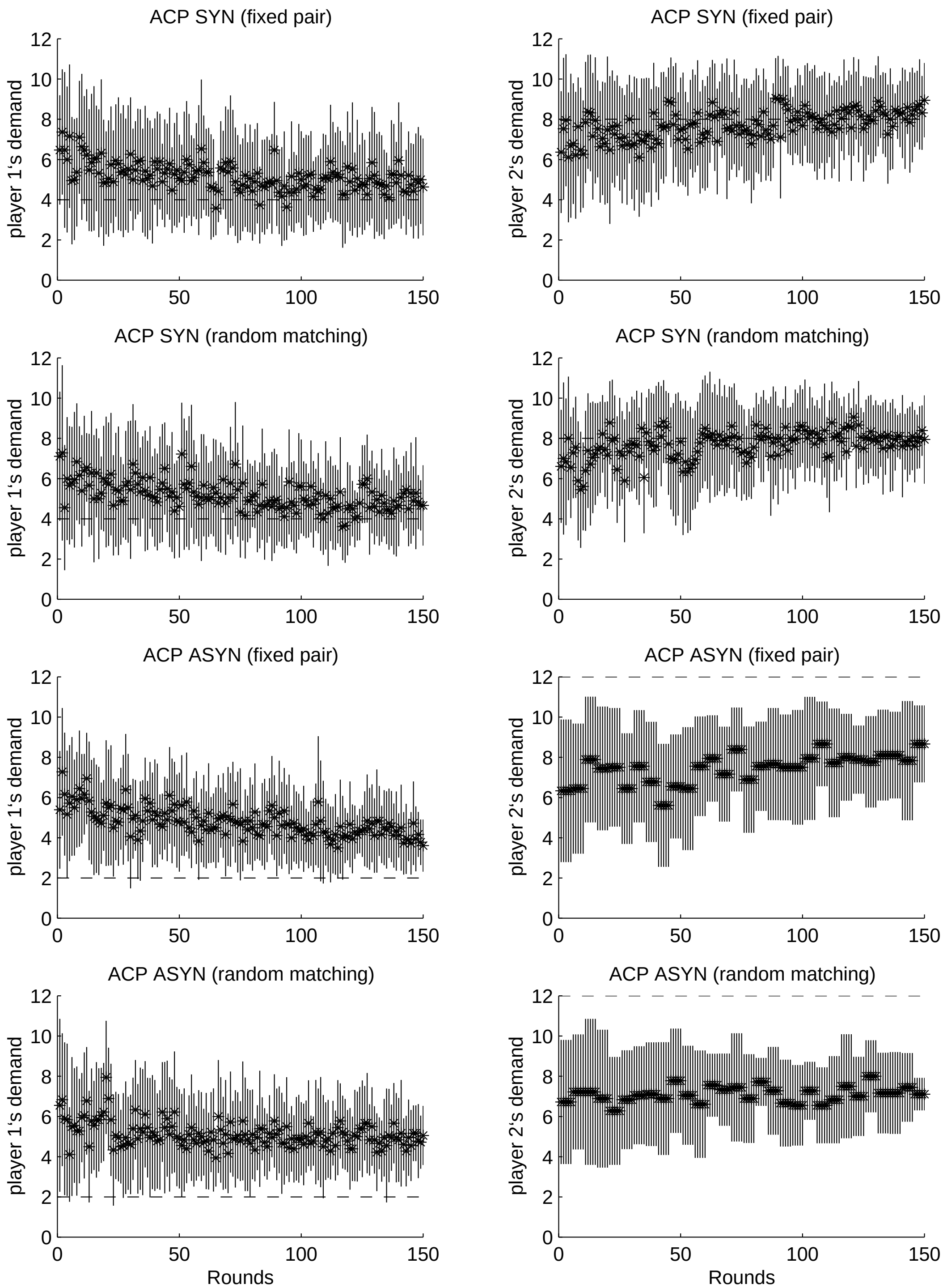


Figure 2. Experimental Data under ACP in Chen (forthcoming).

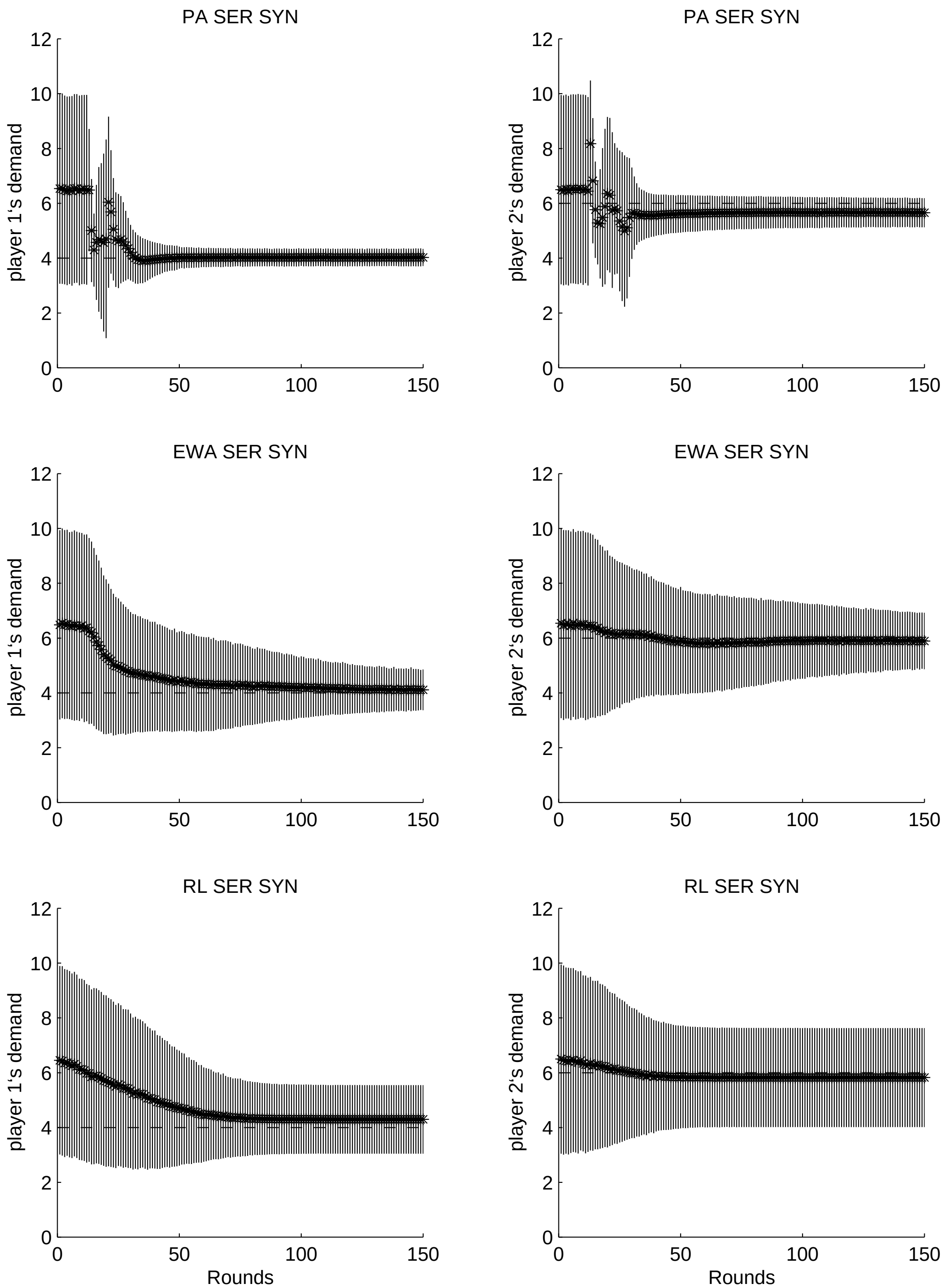


Figure 3. Learning Models under SER SYN (fixed pairs).

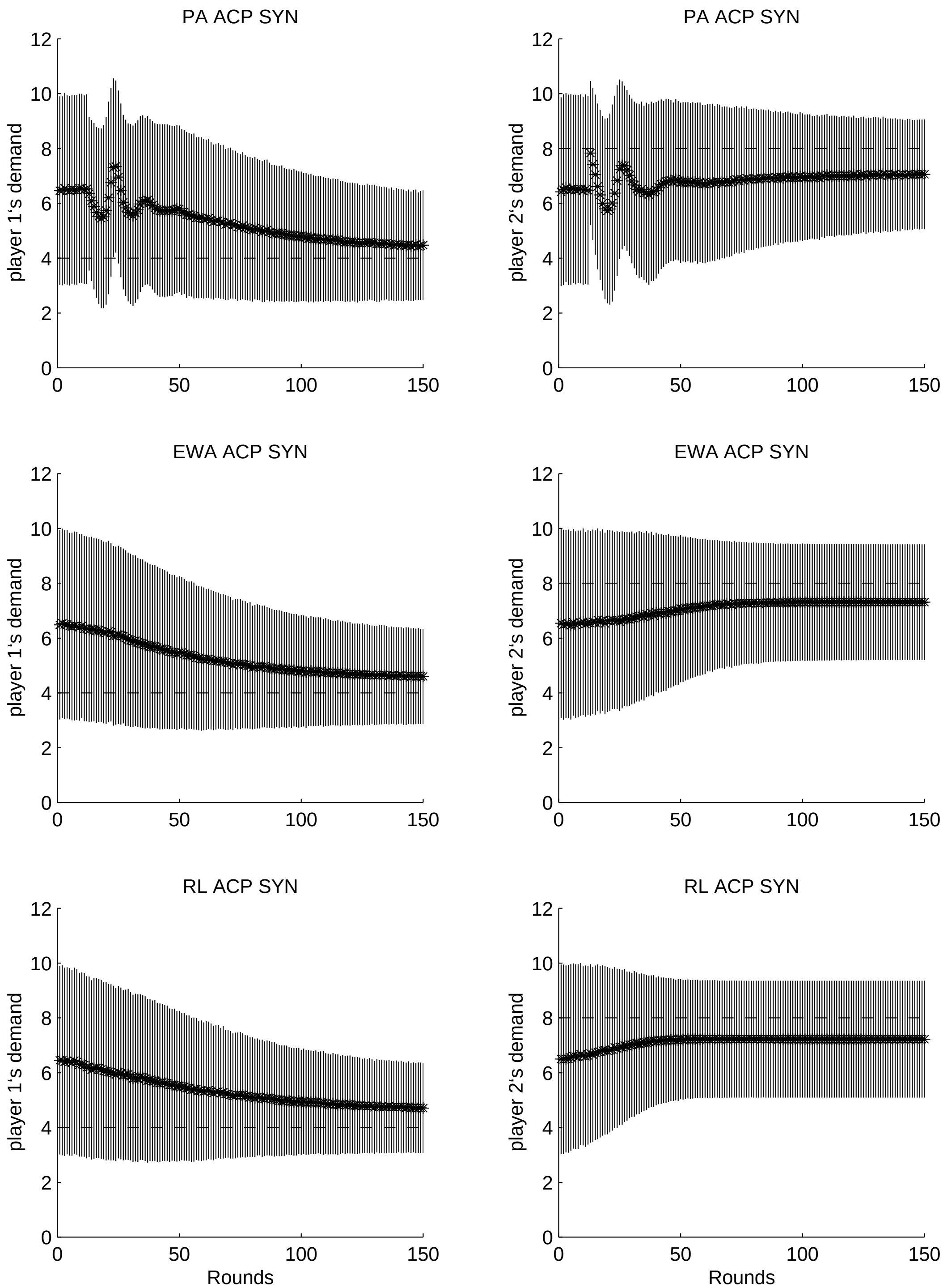


Figure 4. Learning Models under ACP SYN (fixed pairs).

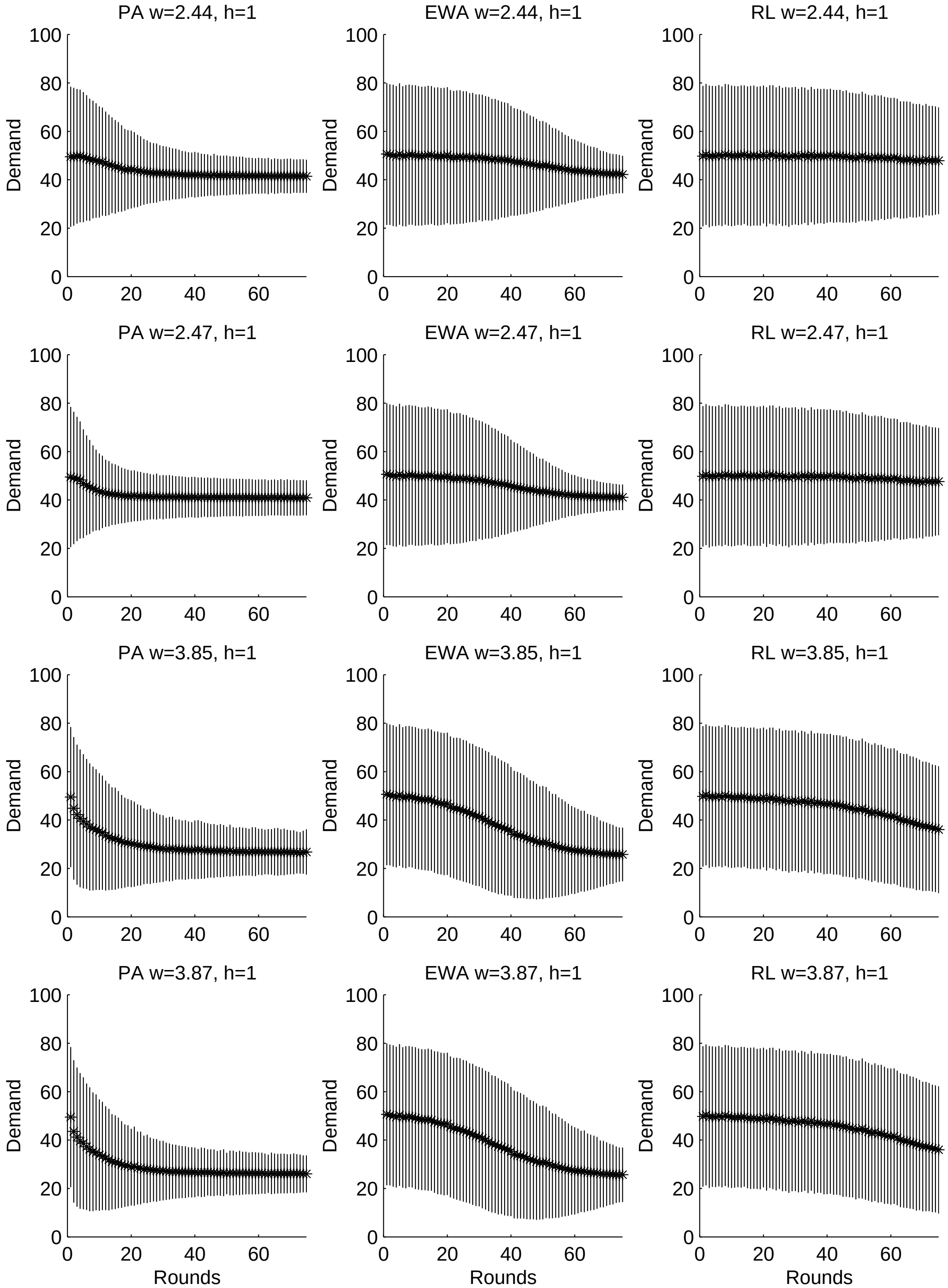


Figure 5. Learning Models Without Similarity Windows in VHBR's Coordination Games.

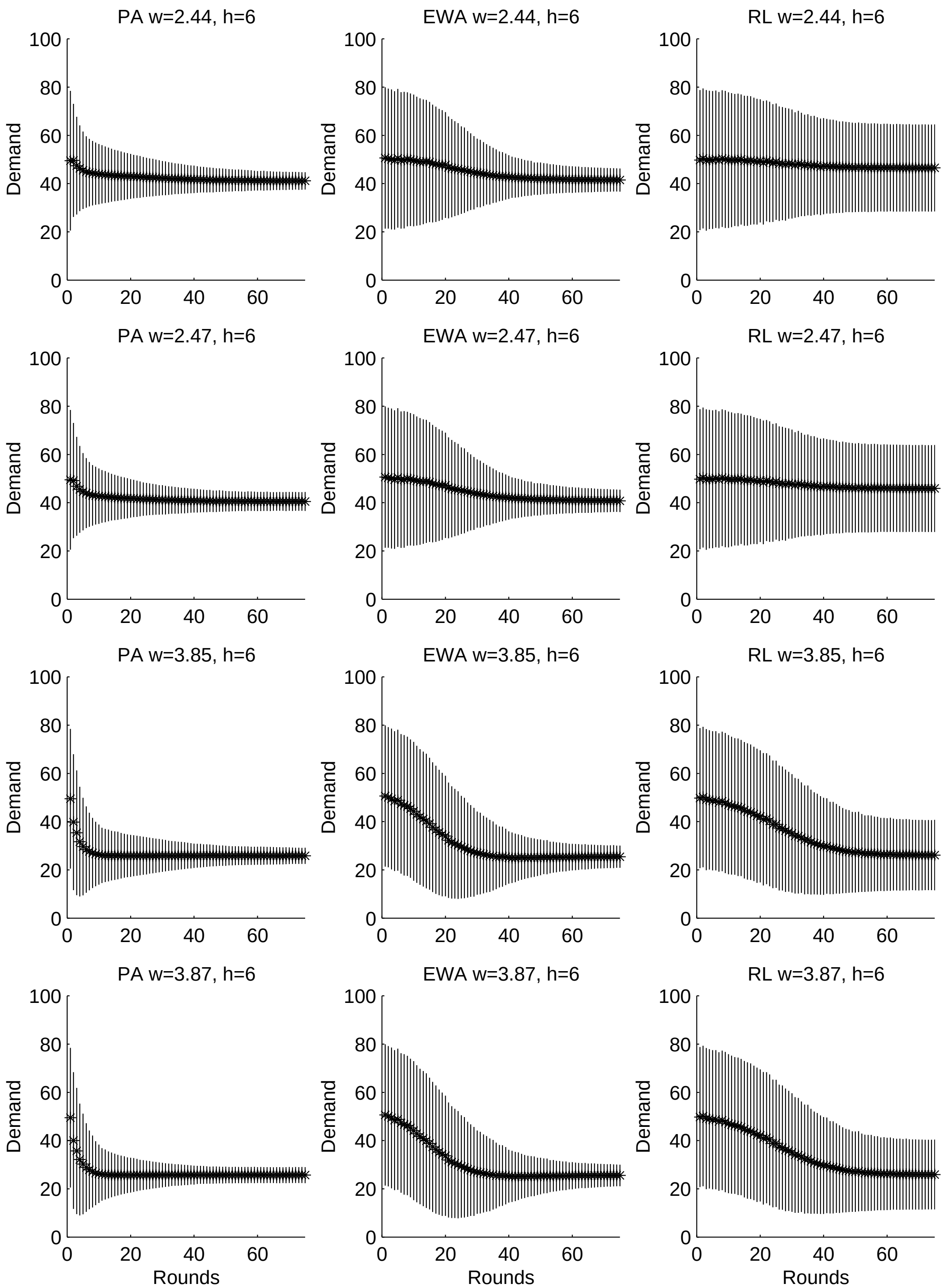


Figure 6. Learning Models With Similarity Windows in VHBR's Coordination Games.