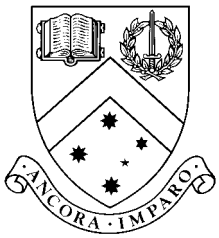
ISSN 1440-771X
ISBN 0 7326 1067 2

M O N A S H U N I V E R S I T Y



AUSTRALIA

Predicting How People Play Games: A Simple Dynamic Model of Choice

Rajiv Sarin and Farshid Vahid

Working Paper 12/99
October 1999

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Predicting How People Play Games: A Simple Dynamic Model of Choice[1]

Rajiv Sarin
Department of Economics
Texas A&M University
College Station, TX 77843
USA

Farshid Vahid
Department of Econometrics
Monash University
Clayton, Victoria 3168
Australia

August 1999

[1]We are grateful to Heather Anderson, Colin Camerer, Larry Samuelson, Klaus Zauner, an associate editor and a referee for comments and encouragement, and to Ido Erev and Barry O'Neill for providing the data.

Abstract

We use the model developed in Sarin and Vahid (1999, *GEB*) to explain the experiments reported in Erev and Roth (1998, *AER*). The model supposes that players maximize subject to their "beliefs" which are non-probabilistic and scalar-valued. They are intended to describe the payoffs the players subjectively assess they will obtain from a strategy. In an earlier paper (Sarin and Vahid (1997) we showed that the model predicted behavior in repeated coordination games remarkably well, and better than equilibrium theory or reinforcement learning models. In this paper we show that the same one-parameter model can also explain behavior in games with a unique mixed strategy Nash equilibrium better than alternative models. Hence, we obtain further support for the simple dynamic model.

Keywords: Payoff assessments without probabilities, games with a unique mixed strategy equilibrium, reinforcement learning.

1 Introduction

In a recent paper, Erev and Roth (1998) show how simple reinforcement learning models can explain behavior in *all* the published experiments involving repeated play of 100 periods or more of games with a unique mixed strategy equilibria. They study not only the *ex-post* descriptive power of the models (by looking at "best fit" parameters), but also their *ex-ante* predictive power (by simulating each experiment based on parameters estimated from other experiments). They show that a one-parameter model robustly outperforms equilibrium predictions in both respects, and that a "generalized" reinforcement learning model outperforms belief based maximization models such as probabilistic fictitious play.

These results are remarkable not only because reinforcement learning models perform better than the equilibrium theory typically used by economists to analyze strategic situations, but also because the reinforcement learning models outperform belief-based models of learning, typically used by economists to describe the dynamic learning behavior of agents in strategic situations. How could a model that does not posit maximizing behavior of any sort and supposes that agents choose among their strategies probabilistically, do better than models which postulate maximizing behavior and that behavior is deterministic which economists typically assume? These results are all the more surprising because the data sets used by Erev and Roth (henceforth, ER) were collected by a variety of other researchers who collected the data under widely different assumptions regarding the information available to the subjects and the manner in which the subjects were paid, if they were paid at all.

In this paper we attempt to explain the same data using a model based explicitly on the maximizing hypothesis, which we introduced in Sarin and Vahid (1999). The model assumes that agents explicitly have "beliefs" that are scalar-valued and non-probabilistic. They represent the subjective assessment of the player regarding the payoff she would obtain from the choice of any strategy at any time. The beliefs are not about how the other players are likely to play but rather they reflect a more immediate concern of the players: the payoffs they "expect" from different strategies. Hence, a player need not know the payoff matrix or even that she is playing a game. Such "beliefs" are intended to describe procedurally rational players who *simplify* the actual, or objective, choice environment they face. Each time the individual is called upon to make a choice she chooses the action that she assesses will give her the highest payoff. The agent, hence, is assumed to be a myopic subjective optimizer. Upon receiving the payoff, which may depend on the state of the world or the choice of other players, the individual updates her assessments.

The Sarin and Vahid (henceforth, SV) model has many similarities with

belief based learning models typically studied by economists (e.g. fictitious play). Like fictitious play,[1] an agent in the SV model is assumed to be myopic – choosing at each time the action she considers to be the best and ignoring the implications of current choices on future choices and payoffs. Like fictitious play, the beliefs of the agent need not be correct and like it the agent updates these beliefs each period according to what she observes. Furthermore, the subjective model of the environment in each is (possibly) mis-specified: In fictitious play this arises because the agent assumes the environment to be stationary, and in SV it arises because the environment is not deterministic. However, unlike fictitious play, the agent's beliefs are not probabilistic. In fictitious play an agent's beliefs concern the probability with which her opponents will play their various strategies. In SV beliefs concern the payoffs the agent "expects"[2] she will receive from the choice of any strategy. In fictitious play the agent is assumed to know the payoff matrix, whereas in SV she need not.

Given the "closeness" of the SV model to those typically studied by economists, and given the ER finding that reinforcement learning models explain the data better than these models, it would be surprising if the SV model would perform better than theirs. This is probably especially so in the class of games with a unique mixed strategy equilibrium. Reinforcement learning models, which postulate that agents choose stochastically, seem more suitable for studying this class of games than the SV model which supposes that choice is deterministic. In coordination games with multiple, symmetric, efficient strict Nash equilibria, we have shown (Sarin and Vahid (1997)) that the SV model explains the data better than equilibrium theory and also considerably better than reinforcement learning models closely related to those studied in ER. It is, therefore, of interest to see how the two models compare in explaining the data on repeated games with unique mixed strategy equilibria.

This paper shows that the SV model explains the data at least as well as the RE model.[3] This is true also when we look at the data period-by-period and compare blocks of several periods with blocks of several periods. In this case, the performance of our model improves over time in comparison with the RE model. This is good news for the traditional economic precepts of maximizing subject to "beliefs" and choosing deterministically. Given that the model does better than the traditional belief-based models considered by ER, it also suggests that the SV model might represent player "beliefs" more appropriately that traditional probability-belief-based models. That is, the manner in which the SV agents are assumed to simplify their choice

[1] See Fudenberg and Levine (1998) for a careful discussion of fictitious play.

[2] or, "assesses" or "anticipates".

[3] Following ER, we refer to the reinforcement learning model they study as the RE model.

environment may be similar to the way in which people actually deal with those environments.

The next section describes the data. Section 3 then presents the SV model and also the ER model. In Section 4 we confront the SV model with the data and present our results and compare them with ER's. We also look here at period-by-period data and see the relative performance of the two models.Section 5 concludes.

2 The Data

The data set is the same as that used by Erev and Roth (1998). The data comes from experiments involving more than 100 repetitions of 12 games with unique mixed strategy equilibria collected under a variety of informational assumptions and payoff structures. We explain these games briefly here. A more detailed summary can be found in Erev and Roth (1996, 1998).

Five of the 12 games were based on Suppes and Atkinson (1960). Four of these five were actually used in experiments conducted by Suppes and Atkinson, which for ease of comparison, we label as in ER: (S&A2, S&A8, S&A3k, S&A3u). All the games are two by two bi-matrix games. Payoffs in these games are binary lotteries. The interaction between the row and column players (players 1 and 2, respectively) determines the probability of the "good" outcome for each player. The following matrices show the probability of the good outcome for each player in these games.

S&A2:	$A2$	$B2$
$A1$	$\left(\frac{1}{3}, \frac{2}{3}\right)$	$(1, 0)$
$B1$	$\left(\frac{1}{2}, \frac{1}{2}\right)$	$\left(\frac{1}{6}, \frac{5}{6}\right)$

S&A8:	$A2$	$B2$
$A1$	$(1, 0)$	$\left(\frac{3}{8}, 1\right)$
$B1$	$\left(0, \frac{5}{8}\right)$	$\left(\frac{5}{8}, \frac{3}{8}\right)$

S&A3:	$A2$	$B2$
$A1$	$(0.3, 0.7)$	$(0.8, 0.2)$
$B1$	$(0.4, 0.6)$	$(0.1, 0.9)$

In S&A2, in equilibrium, player 1 chooses A with probability $\frac{1}{3}$, and player 2 chooses A with probability $\frac{5}{6}$. In the unique mixed strategy equilibrium for S&A8, both players choose A with probability 0.2. S&A3k and S&A3u are two variants of S&A3 with the difference that in S&A3k the payoff matrix is known to the subjects, and in S&A3u the payoff matrix is unknown. In the mixed strategy equilibrium for S&A3, player 1 chooses A with probability $\frac{3}{8}$, and player 2 chooses A with probability $\frac{7}{8}$. S&A2 was played by 20 pairs of players for 200 rounds, and other games were played by 20 pairs of players for 210 rounds. However, round by round data are not available, and only the proportions of A choices for each player in 5 blocks of 40 rounds for S&A2 and 7 blocks of 30 for other experiments are available.

The fifth game based on the Suppes and Atkinson design is labelled S&A3n. It was conducted by Erev and Roth and uses the same payoff matrix as S&A3u, except that correct outcome earns 0.1 Shekel, and incorrect outcome earns 0. The game is played by 10 pairs of players for 500 rounds. Round by round data for this experiment is available, and for compatibility with other models, Erev and Roth (1996) also present this data in frequency of choice A by each player in 10 blocks of 50 rounds.

Game 6 is the game used by Malcolm and Lieberman (1965) and is a standard zero-sum game with payoff matrix,

M&L:	$A2$	$B2$
$A1$	$(3,-3)$	$(-1,1)$
$B1$	$(-9,9)$	$(3,-3)$

where the entries are chips that were converted to money at the conclusion of the experiment. The mixed strategy equilibrium is where player 1 chooses strategy $A1$ with probability 0.75, and player 2 chooses his strategy $A2$ with probability 0.25. This game is played by 9 pairs, who know the payoff matrix, for 200 rounds. Only block averaged data (proportion of choice $A1$ for the row players and proportion of choice $A2$ for the column players) for 8 blocks of 25 rounds are available.

In games 7, 8 and 9 each player has more than 2 available strategies. Game 7, denoted by "On", is a 4×4 zero-sum game from O'Neill (1987) and games 8 and 9, denoted by R&B10 and R&B15, are 5×5 constant-sum games from Rapoport and Boebel (1992). The following matrices show "win" (W) or "lose" (L) outcomes for the row player. Obviously when the row player wins, the column player loses and vice versa.

On	$A2$	$B2$	$C2$	$D2$
$A1$	W	L	L	L
$B1$	L	L	W	W
$C1$	L	W	L	W
$D1$	L	W	W	L

R&B	$A2$	$B2$	$C2$	$D2$	$E2$
$A1$	W	L	L	L	L
$B1$	L	L	W	W	W
$C1$	L	W	L	L	W
$D1$	L	W	L	W	L
$E1$	L	W	W	L	L

In On, win has a payoff of $+5$ and lose has a payoff of -5. In R&B10 and R&B15 the events of win or lose are determined by the matrix R&B, but in R&B10 a W earns $+10$, and an L earns -6, while in R&B15 W earns $+15$, and L earns -1. Equilibrium strategy for On is when both players play A with probability 0.4, and their three other strategies with probability 0.2 each. In both versions of the R&B game, equilibrium play for both players is characterized by the probability vector $\left(\frac{3}{8}, \frac{2}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$. On was played by 25 pairs for 105 rounds, and the frequency of A choices in 7 blocks of 15

are available. R&B games were played by 10 pairs for 120 rounds, and the frequency of A and B choices of each player for 4 blocks of 30 are available.

The other three games are from Ochs (1995), and have the payoff matrix:

OcX	$A2$	$B2$
$A1$	$(\mathrm{X}, 0)$	$(0, 1)$
$B1$	$(0, 1)$	$(1, 0)$

where X=9 in Oc9, X=4 in Oc4 and X=1 in Oc1. Although the games are standard bimatrix games, the experiments were designed to make players think probabilistically. Each player had to announce the proportion of A choices they make in the next 10 rounds. Therefore, even though Oc9 is repeated for 560 rounds, and Oc4 and Oc1 are repeated for 640 rounds, the players only made 56 and 64 rounds of decisions, and only received payoff feedback 56 and 64 times respectively. Eight row players and eight column players are randomly paired in each round of decision making. The blocked averaged data for 7 blocks in Oc9 and 8 blocks in Oc4 and Oc1 of 8 rounds are available.

3 The Model

To simplify notation, we describe the SV model for one of a finite number of players. Suppose that the player has J strategies $S = \{s_1, s_2, ..., s_J\}$. The player associates a subjective assessment with each of her possible strategies. At time $n = 0$, the player's vector of initial assessments is denoted by $u(0) = (u_1(0), ..., u_J(0))$. Her assessment of the payoff she will obtain from choosing strategy s_j at time n, $u_j(n)$, represents her (scalar valued) belief about the payoff she will obtain from the choice of that strategy that time. Her scalar valued beliefs reflect a manner in which she simplifies the decision problem (or game) that she faces. Initial assessments $u(0)$ may have been formed by hearsay, strategy labels, or by similarity of the decision situation to other decision problems that the individual may have faced in the past, as in Gilboa and Schmeidler (1995).

The Sarin and Vahid model is dynamic. It specifies how the individual chooses at each time given her (subjective) assessments at that time, and how these assessments are updated with experience. At each period n the individual is assumed to choose the strategy that she assesses to give the highest payoff. That is, she chooses the strategy that she evaluates to be the best. Implicitly, the individual is myopic, ignoring all the future implications of her current choice on her future choices and payoffs.

The payoff the agent obtains from the choice of any strategy at any time is allowed to be stochastic. The payoff from the choice of s_j when the state of the world is ω is denoted $\pi_j(\omega)$. At the time of making her decision the

agent is not assumed to know the state of the world. When playing a normal form game (repeatedly) with other players, the different possible states of the world correspond to the different possible pure strategy choices of the other players.[4]

If at time n, $u_j(n)$ is the maximum of her assessments, the individual will play s_j. Upon choosing s_j, a state of the world is realized. The decision maker need not observe the state of the world. According to the strategy chosen and the state of the world realized, she receives a payoff. The decision maker then updates her subjective assessments. If she played s_j at time n and the state of the world was ω, she updates her J subjective assessments in the following manner:

$$\begin{aligned} u_j(n+1) &= (1-\lambda)\, u_j(n) + \lambda \pi_j(\omega) \\ u_k(n+1) &= u_k(n) \qquad\qquad \forall k \neq j \end{aligned}$$

where $0 < \lambda < 1$. That is, if the individual chooses s_j and receives $\pi_j(\omega)$, then she updates her subjective assessments about the payoff of s_j by adding a proportion of her surprise (i.e. the difference between the observed payoff and her assessment) to her previous assessment. She does not update her subjective assessments about the payoff that other strategies yield. In the next round, she chooses the strategy which she assesses to be the best, observes a payoff, and adapts her assessments, and so on. Apart from her initial assessments, the only parameter of this model is the "learning parameter" λ.

The model of Sarin and Vahid is in the tradition of models typically postulated by economists. Agents are optimizers. Their optimization is, however, constrained in two ways. Firstly, the agents are not assumed to know the true model of the choice environment they face, and in the light of this incomplete knowledge they simplify the problem they face rather than work with large and extremely cumbersome models. Specifically, they associate with each strategy a scalar rather than a probability distribution (or a probability distribution over a probability distribution) as the traditional Bayesian models would suppose. Secondly, the agents are assumed to be myopic. This can be seen as their response to the incomplete information about their choice environment which they face. It may be the case that because of the uncertainty in the agent's mind about the choice problem the agent behaves myopically.

Next we present the basic (one-parameter) RE reinforcement learning model and contrast it with the SV model.[5] In the RE model of reinforcement learning, the probabilities of different strategies being chosen in period n

[4] We do not specify the distribution according to which the state of the world is chosen. This distribution, in particular, is allowed to be non-stationary.

[5] A more complete presentation of this model, the RE three parameter model, and the ER four parameter fictitious play model can be found in Erev and Roth (1998).

are determined by their "propensities". A player starts with a vector of (positive) initial propensities $(q_1(0), q_2(0), ..., q_J(0))$, which determine the probability of each strategy being chosen in the following way:

$$p_j(0) = \frac{q_j(0)}{\sum_{i=1}^{J} q_i(0)} \qquad j = 1, ..., J$$

If a draw from this distribution leads to choice of s_j, and the state of the world is ω, then the player receives payoff of $\pi_j(\omega)$, and she updates the propensities of different strategies as follows,

$$\begin{aligned} q_j(1) &= q_j(0) + (\pi_j(\omega) - \pi_{\min}) \\ q_k(1) &= q_k(0) \qquad \forall k \neq j \end{aligned}$$

where $\pi_{\min}$ is the minimum of all possible payoffs. Subtracting the minimum payoff ensures that the propensities are updated by a non-negative amount, and hence that they will never become negative. The updated propensities imply the probabilities with which each action will be taken in the next period, and so on. The parameters of this model are the initial propensities (both their relative size across strategies and their absolute magnitude). The one parameter version of this model arises when all the initial propensities are assumed to be the same.

Two important features of reinforcement learning models, and of the RE model in particular, are that choice is described, at each stage, as probabilistic, and that no beliefs are attributed to the agent. A basic principle of this class of models is that actions that result in "good" payoffs are more likely to be played in the future. A specific feature of the (basic) RE model is that all actions that are played are more likely to be played in the future. A more detailed discussion of reinforcement learning models can be found in Börgers and Sarin (1997, 1999) and Roth and Erev (1995).[6]

In the SV model, the parameter λ determines how fast the assessments adapt to the observed payoffs. The larger the λ, the stronger the influence of observed payoffs on assessments. The strength of the influence of the observed payoffs on the assessments is constant at all rounds of play. In the RE model, the larger the q, the smaller the influence of observed payoffs on the propensities would be, and since these propensities are non-decreasing as the play progresses, the influence of observed payoffs on propensities diminishes through time.

4 The Model meets the data

In the first subsection we compare the two models according to the criteria of Erev and Roth. That is we study the *ex post* descriptive power (by

[6] See Camerer and Ho (1999) for a model which "nests" some reinforcement learning models and belief-based models like fictitious play.

looking at "best fit" parameters) and we look at the *ex ante* predictive power (by simulating each experiment based on parameters estimated from other experiments) of the SV model. To study the descriptive power of the SV model we compute the value of λ which minimizes the mean squared deviation (MSD) over all 12 games. We then compare this with the results obtained from the RE model. To study the predictive power of the SV model we compute the value of λ from 11 games and use this to compute the MSD in the remaining game. We then provide a sensitivity analysis to test the robustness of our conclusions. This is done by looking at several alternate specifications of the initial assessments. In the next subsection we check the explanatory power of our model on period by period data when it is available and seems relevant. We explain at that point why we think this is an important test of the explanatory power of the model.

4.1 The Comparison

Our first task is to estimate the learning parameter λ, which is the only parameter of the SV model, based on the observed data. That is, we want to find the value of λ which produces the "best fit" of the SV model to the observed data. The observed data are the block averaged data (i.e. the proportion of A choices in 2×2 games and the proportion of non-symmetrical strategies in the others) for each player in each game, which add up to a total of 180 observations.[7] Following ER we choose the average of the MSD of block averaged data over all 12 games as our measure of fit. Specifically, we find the value of λ that minimizes the MSD between the block averages of data simulated from the SV model and the observed block averaged data, averaged over all 12 games, i.e.,

$$\hat{\lambda} = \arg\min_{\lambda} \left(\frac{1}{12} \sum_{i=1}^{12} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \left(\bar{y}_{ij}^{s}(\lambda) - \bar{y}_{ij}^{o} \right)^2 \right) \right)$$

where $\bar{y}_{ij}^{o}$ is the j-th observed block averaged data in game i, $\bar{y}_{ij}^{s}(\lambda)$ is the corresponding simulated block averaged data for game i, and n_i is the number of blocks in game i. Since λ is in $[0,1]$, grid search is the most efficient optimization method for this problem. To obtain the simulated block averages, each game is simulated 200 times, and in the simulated games, the players are paired the same way as in the actual experiments, and each game is repeated the same number of rounds as in the actual experiments. The length of each game is then divided to as many blocks

[7] There are 5×2 observations in S&A2, 7×2 observations in each of S&A8, S&A3u and S&A3k, 10×2 observations in S&A3n, 8×2 observations in M&L, 7×2 observations in On, $4 \times 2 \times 2$ (4 blocks of two non-symmetric choices for two players) observations in each of R&B10 and R&B15, 7×2 observations in Oc9, and 8×2 observations in each of Oc4 and Oc1.

as the experimental data have been, and then the choice indicators in each block are averaged to yield proportions of each choice for each block.

The initial assessments for each game are drawn uniformly from $[U_{min},\ U_{max}]$, where U_{min} is the minimum payoff a player may obtain in the game and U_{max} is the maximum payoff a player may obtain. Even though in the games where payoffs are known one can think of more plausible methods of assigning initial assessments, we chose this method because it leads to uniform play (i.e. all strategies are equally likely) in the first round of play, which is the assumption made by Erev and Roth.[8]

Some further choices had to be made to complete the correspondence between the SV model and the RE model for the aforementioned experiments. We had to make choices about S&A games in which there were no monetary payoffs, and Oc games in which the players did not actually play the game characterized by the Oc payoff matrix stated in the previous section, but rather they chose how many choice A's they were going to make in 10 simultaneous rounds of that game. For the S&A games we assigned a payoff of 1 to a "correct" choice, and 0 to an "incorrect" choice. This choice is arbitrary, but because we choose initial conditions from $[U_{min}, U_{max}]$ it doesn't matter what we assign to good and bad payoffs. For the Oc games, we recognize that each player does not just have two choices consisting of playing A or not playing A, but rather they have 11 choices of playing A 0 to 10 times in the next 10 rounds. To account for this, initial assessments are drawn uniformly for each of these 11 choices. The optimal choice is made based on them, and assessments for these 11 choices are updated. The results of the basic comparison between the (one-parameter) SV model and the RE models are summarized in the Table 1.

(Table 1 about here)

The value of the parameter λ which minimizes the MSD of the simulated and observed data is $\hat{\lambda} = 0.010$, and the overall minimum MSD is 0.92. Although the minimization is done over all games together, game by game MSD scores are also provided in the first row of Table 1. For comparison purposes, the second to the fifth row of this table quotes the MSD scores for the best fitted one parameter RE model, for the best fitted 3 parameter RE model, for the best fitted 4 parameter fictitious play model,[9] and for

[8] Note that, as initial assessments are drawn from $[U_{min}, U_{max}]$, when payoffs undergo an affine transformation the initial assessments are appropriately modified, and choice behavior remains unaffected. We are grateful to Ido Erev for suggesting this choice of initial assessments.

[9] For the sake of brevity, we do not discuss the three parameter RE model of reinforcement learning or the four parameter fictitious play model. The reader should consult the ER paper for these models.

the equilibrium play, are all taken from Table 1 of ER. As can be seen from this table, the MSD score of the SV model is lower than the MSD score of the basic reinforcement model of RE (i.e. RE (1 par.)). The games for which the SV model fits worse than the RE model are mainly the S&A game experiments, in which there were no monetary payoffs.

The sixth row of Table 1 presents the MSD score of each of the 12 games, when the data for each game is simulated based on the parameters that best fit the other 11 games. We perform these calculations, as do Erev and Roth, to test the predictive power of the model. The overall mean of the MSD for the 12 games with the SV model is 1.01 which is similar to the 1.02 obtained by ER (see their Table 1) for the basic RE model.

The last two rows of Table 1 report the MSD fit of the SV model if a different learning parameter is chosen for each game, and the optimal value of this parameter for each game. The overall MSD score of the SV model optimized by game is 0.43, which is significantly smaller than that of the basic RE model by game (0.68), and it is slightly above that of the 3 parameter RE model by game (0.35).[10]

In order to ensure that the favorable results of the SV model is not a consequence of a lucky choice of initial assessments, we examine several different sets of initial assessments. We re-estimate the model with five different sets of initial assessments: (i) a truncated normal distribution with mean equal to the average payoff and standard deviation equal to the standard deviation of the payoffs but truncated so that all initials are in the range of possible payoffs; (ii) a normal distribution with mean equal to the average payoff and standard deviation equal to the standard deviation of the payoffs, (iii) a normal distribution with mean equal to the average payoff and standard deviation equal to half of the standard distribution of the payoffs, (iv) a normal distribution with mean equal to the average payoff and standard deviation equal to twice the standard deviation of the payoffs, and (v) a uniform with range from the maxmin payoff (which is the payoffs the subjects could over secured themselves in the game) and the maximum payoff.

Table 2 shows the estimated parameters $\left(\hat{\lambda}\right)$ and the overall mean deviation scores for the above five sets of initial assessments. As it can be seen from this table, all these alternate initial assessments result in MSD values which are lower than those we obtain with our basic set of initial assessments. This suggests that the minimized value of the MSD over all games is not too sensitive to reasonable alternative distributions of initial assessments.

(Table 2 about here)

[10] See Table 1 of ER for game by game MSD scores of variants of the RE model.

4.2 An Extension

We take the analyses reported above as an indication that the SV model is a useful model of learning and choice in a repeated game setup when games have a unique mixed strategy equilibrium. However, we have reservations about comparing learning models based on their blocked averaged MSD scores. Firstly, the MSD measure has the undesirable mathematical property that it is not invariant to the choice of the number of blocks. That is, it is possible for model X to have smaller MSD score than model Y if the 80 rounds of play are divided to 8 blocks of 10 rounds, but to have larger MSD score than model Y if the 80 rounds of play are divided to 4 blocks of 20 rounds. This shortcoming can be remedied by using an alternative measure which would be invariant to merging or splitting the blocks. The literature on invariant measures for model forecast comparison in econometrics can lead us to define appropriate measures for this purpose (e.g. Clements and Hendry (1994)).

Secondly, if we suppose that the payoff distribution was stationary, we could derive the implied stochastic processes which block averaged choices by each model would follow. This would allow us to design statistical tests to refute (or not) each model. However, in a repeated game situation, the payoff distribution is evolving endogenously over time, and is non-stationary. Deriving the properties of block averages in this situation would be extremely difficult.[11] Hence, we need to think of other ways of testing the compatibility of alternative models with the data. Our suggestion is to use the fact that there are repeated observations of each game in each experiment (usually games are played by 10 to 25 pairs of players). Hence we have an empirical distribution of choice in each round of play, and we also have a model-implied distribution of choice in each round for each model based on simulated data. We can then test the closeness of the two distributions at each round of play for each model. The results of these tests will reveal a lot more information than a summary statistic such as the MSD score can. They can tell us which model fits the initial play better, and more importantly, they can tell us which model "learns" better, i.e. processes the observed information similar to how the experimental subjects do.

Round by round data are available for the O'Neill experiment (On), the Erev and Roth new experiment based on the Suppes and Atkinson design (S&A3n), and the Ochs experiments (Oc9, Oc4 and Oc1). From Table 1, we can see that the SV model has smaller MSD score in the O'Neill and Ochs games (Oc9 and Oc4) than the RE model, and a worse MSD score in S&A3n than the RE model. We only use O'Neill and Erev-Roth data to compare

[11] Note that the cross section averages over all individuals at one time, and time series averages for each individual over all rounds in a block are quite different random variables in these models because of learning and feed-back in choices over time. The block averages are a mixture of these two.

the round by round performance of the RE and the SV models against the data. We do not use Och's data because of the complicated design of those experiments (recall that at each stage each player was asked to announce the proportion of choice A in the next 10 rounds, and that the players were randomly matched at each stage).

We use the Pearson's statistic (see Kendall, Stuart and Ord, 1991, chapter 30) to compare the observed and theoretical distributions at each round. In the O'Neill experiment, 25 pairs of players had to choose one of their four available strategies at each round for 105 repetitions. Let n_{ij} denote the number of row players who chose alternative $i \in \{1,2,3,4\}$ at round j in the actual experiment. Also let p_{ij}^s be the proportion of the row players who choose alternative i at round j according to one of the theoretical models, derived through many (here 1000) simulations. Then the Pearson's statistic for testing the null hypothesis that the theoretical and the observed distributions are the same is:

$$X^2 = \sum_{i=1}^{4} \frac{\left(n_{ij} - 25p_{ij}^s\right)^2}{25p_{ij}^s}$$

which has a χ^2 distribution with 3 degrees of freedom.

Table 3 shows the number of times the hypothesis of equality of the observed and the theoretical distributions was rejected at the 5% significance level,[12] for the row and the column players when the theoretical distributions are simulated from (i) the SV model that is fitted to block averaged data for O'Neill game only ("SV by game"), (ii) the SV model which is fitted to block averaged data for all 12 games ("SV over all"), (iii) the RE model that is fitted to all 12 games ("RE over all"), and (iv) the equilibrium play ("EQ", the equilibrium play probabilities for this game are $(0.4, 0.2, 0.2, 0.2)$ for both players). It can be seen from this table that the SV model which is fitted exclusively to block averaged data from the O'Neill experiment, performs well in learning from the observed information. In particular, the SV model gets closer to the empirical distribution as the game progresses.

(Table 3 about here)

In the Erev-Roth new experiment with S&A3 design and monetary payoff of 0.1 shekels for a "correct" choice and zero for an "incorrect" choice, the

[12] The level of significance of each individual test is 5%. However, the tests statistics for different rounds are not independent, and therefore the level of significance of the entire procedure might be different from 5%.

observed data come from 9 pairs[13] of players choosing one of their two choices at each round for 500 rounds. In this case, using the same notation as above, the Pearson's statistic for testing the null hypothesis that the theoretical and the observed distributions are the same is:

$$X^2 = \sum_{i=1}^{2} \frac{\left(n_{ij} - 9p_{ij}^s\right)^2}{9p_{ij}^s}$$

which has a χ^2 distribution with 1 degree of freedom. The simulated probabilities for each model were calculated from 1000 replications.

Table 4 shows the number of times that the SV model fitted to block averaged data of the S&A3n experiment and the SV and RE models that are fitted to block averaged data for all games, and the equilibrium play, are rejected over all 500 rounds, and in the first, second, third, fourth and fifth 100 rounds of play. The equilibrium play in S&A3 game is $\left(\frac{3}{8}, \frac{5}{8}\right)$ for the row player, and $\left(\frac{7}{8}, \frac{1}{8}\right)$ for the column player. With the payoff matrix being so complex and neither player knowing it, there is no reason to expect that equilibrium play has any significance in this setup. Both the MSD score and the round by round statistics show that it does not have any descriptive power for the experimental data. Even though the SV model has a higher MSD score than the RE model in Table 1 (0.85 versus 0.57), Table 4 shows that the SV model fits the round by round experimental data as well as the RE model. Both models are rejected about 10% of the times in all 500 rounds for each player. The SV model that is fitted to the S&A3n data exclusively, is rejected less than 5% of the time for each player.

(Table 4 about here)

5 Conclusion

We have shown that the one parameter simple dynamic model introduced in Sarin and Vahid (1999) can explain the experimental data in games with unique mixed strategy equilibria at least as well as the reinforcement learning model of Roth and Erev (1995). We obtained this result both in terms of *ex post* descriptive power and in terms of *ex ante* predictive power of the model. Our conclusions did not change when we evaluated the data

[13] Due to a technical problem in the file transfer, the data of one of the original pairs 10 pairs was not used.

at a more disaggregated, period by period, level. These results are good news for the traditional economic principles of choosing deterministically and with a view to maximizing some objective. The results suggest that models, which incorporate the manner in which individuals simplify their objective environment, might be adequate in explaining large bodies of data reasonably well.

Some caveats regarding these conclusions are, however, in order. First, even while we have used *all* the data that Erev and Roth could find with 100 of more repetitions involving repeated play of games with a unique mixed strategy equilibria, more data would be required for firmer conclusions. Our earlier analysis (Sarin and Vahid (1997)), using data from repeated coordination games, reveals that some of these conclusions are robust to different games. Second, our analysis did find that probabilistic choice models might be useful in explaining data in some games. Further work is required in determining the games and conditions under which probabilistic choice models might perform well. This paper suggests a question regarding how different the predictions based on stochastic or deterministic choice models may be. In future work we plan to conduct an experiment which would differentiate between the deterministic and stochastic choice models.

Our analysis, like that of Erev and Roth, has largely been concerned with average choice frequencies. We could, however, also be interested in other aspects of the data. One such concern could be to try and explain individual learning curves as opposed aggregate learning curves. A first place to study individual learning curves would be in decision theoretic data involved two-armed bandit models. Bereby-Meyer and Erev (1998) have recently gathered and studied such data. Another area of research that we believe is important is the careful econometric evaluation of alternate learning models. This would help in further understanding alternate models, and distinguishing their key features.

REFERENCES

1. Bereby-Meyer, Y. and I. Erev (1998): "On learning to become a successful loser: A comparison of alternative abstractions of learning processes in the loss domain," *Journal of Mathematical Psychology*, 42, 266-286.

2. Börgers, T. and R. Sarin (1997): "Learning through reinforcement and replicator dynamics," *Journal of Economic Theory*, 77, 1-14.

3. Börgers, T. and R. Sarin (1999): "Naive reinforcement learning with endogenous aspirations," *International Economic Review*, forthcoming.

4. Camerer, C. and T. Ho (1999): "Experience weighted attraction learning in normal-form games," *Econometrica*, 67, 827-874.

5. Clements, M. and D. Hendry (1994): "Towards a Theory of Economic Forecasting," in C. Hargreaves, *Nonstationary Time Series Analysis of Co-integration*, Oxford University Press.

6. Erev, I. and A. Roth (1996): "On the need for low rationality, cognitive game theory: Reinforcement learning in experimental games with unique, mixed strategy equilibrium," mimeo, University of Pittsburgh

7. Erev, I. and A. Roth (1998): "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibrium," *American Economic Review*, 88, 848-881.

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

9. Gilboa, I. and D. Schmeidler (1995): "Case based decision theory," *Quarterly Journal of Economics*, 110, 605-639.

10. Kendall, M., A. Stuart and J. Ord (1991): *Advanced Theory of Statistics, Vol. 2* (5th Edition)

11. Malcolm, D. and B. Liebermann (1965): "The behavior of responsive individuals playing a two-person, zero-sum game requiring the use of mixed strategies," *Psychonomic Science*, 373-374.

12. Ochs, J. (1995): "Simple games with unique mixed-strategy equilibrium: An experimental study," *Games and Economic Behavior*, 10, 202-217.

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

14. Rapoport, A. and R. Boebel (1992): "Mixed strategies in strictly competitive games: A further test of the minmax hypothesis," *Games and Economic Behavior*, 4, 261-283.

15. Roth, A. and I. Erev (1995): "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate run, *Games and Economic Behavior*, 6, 164-212.

16. Sarin, R. and F. Vahid (1999): "Payoff assessments without probabilities: A simple dynamic model of choice," *Games and Economic Behavior*, 28, 294-309.

17. Sarin, R. and F. Vahid (1997): "Payoff assessments without probabilities: Incorporating "similarity" among strategies," mimeo, Texas A&M University and Monash University.

18. Suppes, P. and R. Atkinson (1960): *Markov Learning Models for Multiperson Interactions*, Stanford University Press.

Table 1: MSD scores ($\times 100$) for the SV model and various versions of the RE model

Game $\rightarrow$	S&A	S&A	S&A	S&A	S&A	M&L	On	R&B	R&B	Oc	Oc	Oc	Mean
Model $\downarrow$	8	2	3u	3k	3n			10	15	9	4	1	overall
SV(1 par.)	0.26	0.78	0.37	0.64	0.85	0.52	1.69	0.89	0.87	2.53	1.08	0.53	0.92
RE(1 par.)	0.16	0.30	0.31	0.11	0.57	2.27	1.81	0.73	0.98	2.71	1.54	0.48	1.00
RE(3 par.)	0.38	0.18	0.12	0.07	0.31	1.24	0.72	0.33	0.65	1.54	1.09	0.48	0.59
FP(4 par.)	0.34	0.20	0.16	0.09	0.37	1.26	1.05	0.44	0.71	2.04	1.48	0.42	0.71
Equilibrium*	6.91	7.18	7.30	7.60	6.12	2.11	0.14	0.48	1.06	2.24	1.37	0.44	3.57
SV prediction	0.31	1.05	0.45	0.65	1.28	0.71	1.69	0.89	0.87	2.53	1.08	0.61	1.01
SV by game	0.21	0.31	0.37	0.59	0.41	0.51	0.29	0.15	0.28	1.16	0.55	0.34	0.43
(12 par.) $\hat{\lambda}_i$	0.012	0.004	0.009	0.015	0.004	0.011	0.979	0.939	0.904	0.110	0.059	0.950	

*There is a slight discrepancy between the MSD scores of the equilibrium play that we have calculated, and those reported in Table 1 of Erev and Roth (1998). We have not found out the source of this discrepancy, yet. The differences though are very minor.

Table 2: Estimation results for alternative sets of initial assessments

Initial Assessments	$\hat{\lambda}$	Overall MSD
$TN(\bar{\pi}, s_\pi, \pi_{\min}, \pi_{\max})$	0.010	0.83
$N(\bar{\pi}, s_\pi)$	0.025	0.77
$N(\bar{\pi}, 0.5 * s_\pi)$	0.007	0.91
$N(\bar{\pi}, 2 * s_\pi)$	0.120	0.71
$U(\pi_{\max\min}, \pi_{\max})$	0.009	0.83

Table 3: Rejection frequencies of the theoretical models when confronted with the O'Neil data

Model	Row player	Column player
	Over all 105 rounds	
SV by game	11	8
SV over all	32	10
RE over all	30	10
EQ	9	1
	Initial 35 rounds	
SV by game	7	2
SV over all	11	9
RE over all	9	9
EQ	7	1
	Middle 35 rounds	
SV by game	3	4
SV over all	9	1
RE over all	8	1
EQ	1	0
	Final 35 rounds	
SV by game	1	2
SV over all	12	0
RE over all	13	0
EQ	1	0

Table 4: Rejection frequencies of the theoretical models when confronted with the Erev-Roth new experimental data (S&A3n)

Model	Row player	Column player
	Over all 500 rounds	
SV by game	24	23
SV over all	42	48
RE over all	52	46
EQ	153	219
	First 100 rounds	
SV by game	9	7
SV over all	5	6
RE over all	9	7
EQ	23	57
	Second 100 rounds	
SV by game	2	6
SV over all	1	2
RE over all	3	9
EQ	13	39
	Third 100 rounds	
SV by game	4	3
SV over all	5	11
RE over all	10	7
EQ	32	50
	Fourth 100 rounds	
SV by game	8	2
SV over all	15	24
RE over all	14	7
EQ	39	50
	Fifth 100 rounds	
SV by game	1	5
SV over all	16	5
RE over all	16	16
EQ	46	23