

4 On the Role of Reinforcement Learning in Experimental Games: The Cognitive Game-Theoretic Approach

Ido Erev

Technion – Israel Institute of Technology

Alvin E. Roth

University of Pittsburgh

Experimental investigation of choice behavior in repeated games reveals that in many settings the effect of experience can be predicted by simple reinforcement learning models. These models, which feature a slow adjustment process in response to accumulated reinforcements, appear to capture behavior when it is both consistent and inconsistent with equilibrium predictions (see e.g., Bornstein, Erev & Goren, 1994; Camerer & Ho, chap. 3, this volume; Erev and Rapoport, 1998; Erev & Roth, 1998; Mookherjee & Sopher, 1997; Rapoport, Rapoport, Abraham, & Olson, 1997; Rapoport, Seal, Erev, & Sundali, 1998; Roth & Erev 1995; Slonim & Roth, 1998; Tang, 1996).

Yet examination of the experimental psychology and gaming literatures also reveals robust learning phenomena that cannot be explained by simple reinforcement learning models. These phenomena include sequential dependencies (Rapoport & Budescu, 1992; Tolman, 1925), reciprocity (Rapoport & Channah, 1965), transfer (e.g., Rapoport et al., 1998), expectation-based learning (Camerer & Ho, chap. 3, this volume), imitation (Bandura, 1969), and direction learning (Selten & Buchta, chap. 5, this volume).

Three classes of explanations can be provided to account for the apparent contradiction between the good descriptive power (in some settings) and the clear violations (in other settings) of simple reinforcement learning models. First, in line with the general approaches proposed by Camerer (1990), Camerer and Ho (chap. 3, this volume), Cooper and Feltovich (1996), Mookherjee and Sopher (1997), and Selten and Buchta (chap. 5, this volume), it can be argued

that choice behavior and learning are game-specific. Thus, it is possible that in some games learning is best approximated by a reinforcement learning process, whereas other processes underlie learning in other games.

According to a second class of explanations reinforcement learning models do not capture a robust property of human behavior. Rather, human learning is better approximated by a different general learning (or other) rule whose predictions are similar to the predictions of reinforcement learning models in some games. An argument of this type is made in Fudenberg and Levine's (1997) book on learning theory and by Cheung and Friedman (1996).

Note that the first of the two classes of explanations presented, and to some extent the second, are rather pessimistic. The first implies that a general model of learning could be developed, if at all, only after we understand learning in a very large set of specific games (this model would predict which learning rule would be used in a new game). And the second suggests that the knowledge accumulated thus far in experimental research is of little value; we are yet to find a general principle of learning (unless one of the new proposals should prove surprisingly robust). The third class of explanations, proposed in our recent work (Erev & Roth, 1998; Roth & Erev, 1995) and referred to as cognitive game theory, is more optimistic. This approach conjectures that learning can always be approximated by reinforcement learning models in which people learn among cognitive strategies. Thus, whenever high level cognitive strategies play an important role, reinforcement models that ignore these strategies fail. This approach predicts that when the relevant cognitive strategies are understood (and much of the research in cognitive psychology focuses on that goal), the effect of experience can be predicted by a general model.

The main goal of this chapter is to review the evidence that suggests the role of reinforcement learning is best understood within the cognitive game theory framework. This chapter: (a) describes Roth and Erev's (1995) reinforcement learning model; (b) reviews results that demonstrate that this simple model provides a good approximation of behavior in repeated games in which players cannot reciprocate; (c) summarizes six violations of this model (and of the simple reinforcement learning approach); (d) presents the cognitive game-theoretic explanation of these violations, and discusses the value and limitations of this approach; and (e) presents conclusions and directions for future research.

ROTH AND EREV'S REINFORCEMENT LEARNING MODEL

In Roth and Erev (1995), we proposed a reinforcement learning model built on a linear quantification of Thorndike's (1898) law of effect (similar quantifications were suggested by Bush & Mosteller, 1955; Harley, 1981; Herrnstein, 1970; and Luce, 1959). To the basic quantification we added three additional important characteristics of human and animal learning: generalization (and

experimentation), recency, and effect of reference points (Erev & Roth, 1996b).

Basic Assumptions and a One-Parameter Model

The family of models we considered can be summarized by four main assumptions. For expository purposes these assumptions are presented with the specific quantification assumed by a basic one-parameter model.

A1. Initial propensities. At time $t = 1$ (before any experience has been acquired) each Player n has an initial propensity to play k th pure strategy, given by some nonnegative number $q_{nk}(1)$. In the basic model, each player will be assumed to have equal initial propensities for each pure strategy, that is, for each Player n ,

$$q_{nk}(1) = q_{nj}(1) \text{ for all pure strategies } k, j. \quad (1)$$

A2. Reinforcement function. The reinforcement of receiving a payoff x is given by an increasing function $R(x)$. In the basic model the reinforcement function is set to

$$R(x) = x - x_{\min} \quad (2)$$

where x_{\min} is the smallest possible payoff.

A3. Updating of propensities. If Player n plays k th pure strategy at time t and receives a reinforcement of $R(x)$, then the propensity to play strategy j is updated as a function of $R(x)$. The basic model assumes a linear function,

$$q_{nj}(t+1) = \begin{cases} q_{nj}(t) + R(x) & \text{if } j = k \\ q_{nj}(t) & \text{otherwise} \end{cases} \quad (3)$$

A4. Probabilistic choice rule. Following Luce (1959) the probability $P_{nk}(t)$ that Player n plays his k th pure strategy at time t is

$$P_{nk}(t) = \frac{q_{nk}(t)}{\sum_j q_{nj}(t)} \quad (4)$$

where the sum is over all of Player n 's pure strategies j .

Note that the model satisfies the law of effect (Thorndike, 1898) and the power law of practice (Blackburn, 1936). Pure strategies that have been played

and have met with success tend to be played with greater frequency than those that have met with less success, and the learning curve will be steeper in early periods and flatter later (because nonnegative reinforcements imply $\sum q_{nj}(t)$ is an increasing function of t , so a reinforcement of $R(x)$ from playing pure strategy k at time t has a bigger effect on $P_{nk}(t)$ when t is small than when t is large).

The Single Parameter of the Basic Model. It follows from the probabilistic choice rule (Eq. 4) and our assumption that each player's initial propensities are all equal that at the initial period of the game each player chooses each strategy with equal probability. However we have not made any assumption that fixes the sum of the initial propensities, which appears in the denominator of Eq. 4, and therefore influences the rate of change of choice probabilities, that is, the speed of learning (which is also influenced by the size of the rewards). The basic model's sole parameter, $s(t)$, which we call the *strength* of the initial propensities, is introduced to determine the ratio of these two determinants of the learning speed. Let X_n be the average absolute payoff for Player n in the game. The initial strength parameter for Player n is defined as $s_n(t) = \sum q_{nj}(1)X_n$, and we assume that this is a constant for all players, that is, $s_n(t) = s(t) > 0$ for all Players n .

Note that this definition and the probabilistic choice rule yield the initial propensities $q_{nj}(1) = P_{nj}(1) s(1) X_n$, where $P_{nj}(1)$, the initial choice probability is given by $P_{nj}(1) = 1/M_n$, where M_n is the number of Player n 's pure strategies. Thus the initial propensities are determined by the observable features of the game and by the strength parameter $s(1)$.

A Three-Parameter Model

It is easy to see that the basic model can come to adjust too slowly. That is, if $s(1)$ is low it predicts fast initial learning, but extremely slow learning in the longer term (when the propensities are large relative to the obtained reinforcements). To address this problem we (Roth & Erev, 1995) introduced responsiveness to the model by adding two weaker psychological assumptions: experimentation and a recency effect. The first of these can be viewed as an extension of the law of effect:

Experimentation (or Generalization): Not only are choices that were successful in the past more likely to be employed in the future, but similar choices will be employed more often as well, and players will not (quickly) become locked in to one choice in exclusion of all others.

The second additional feature of individual learning modeled in Roth and

Erev (1995), can be viewed as an interaction between the law of effect and the power law of practice.

Recency. Recent experience may play a larger role than past experience in determining behavior.

In Roth and Erev (1995) we called this "forgetting." These two assumptions were quantified in Roth and Erev by the following modification of Eq. 3 (assumption A3), the updating function:

$$q_{nj}(t+1) = (1-\phi)q_{nj}(t) + E_{nk}(j, R(x)) \tag{5}$$

In Eq. 5, ϕ is a forgetting (or recency) parameter that slowly reduces the importance of past experience, and E is a function that determines how the experience of playing strategy k and receiving the reward $R(x)$ is generalized to update each strategy j .

Experimental investigation of generalization suggests that strategies that subjects find similar to the selected strategy will be affected by the reinforcement. Brown, Clark, & Stein (1958) observed a normal generalization distribution. In games in which similarity of strategies can be linearly ordered (such as those studied in Roth & Erev, 1995) we chose a three step function to approximate the generalization function, as follows:

$$E_{nk}(j, R(x)) = \begin{cases} R(x)(1-\epsilon) & \text{if } j = k \\ R(x)\epsilon/2 & \text{if } j = k-1, \text{ or } j = k+1 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where ϵ is an experimentation/generalization parameter. For games when only two strategies are considered, or when the $M_n \geq 2$ strategies do not have an apparent linear order, the generalization function is reduced to a two step function:

$$E_{nk}(j, R(x)) = \begin{cases} R(x)(1-\epsilon) & \text{if } j = k \\ \frac{R(x)\epsilon}{(M_n-1)} & \text{otherwise} \end{cases} \tag{7}$$

Another way to think of these two functions is that when the strategy sets allow similarity judgments to be made, players will generalize their most recent experience in a way that leads to experimentation among the most similar strategies. When no similarity judgments can be made, players simply retain some propensity to experiment among all strategies.

Parameters. The model has three parameters: the strength parameter $s(1)$ (as in the basic model) and the experimentation and forgetting parameters ϵ and ϕ .

SUPPORTIVE EVIDENCE

Analysis of Published Data Sets

To evaluate the models just described, Erev and Roth (1998) assembled and analyzed a data set consisting of all experiments we could locate involving play of 100 periods or more of games with a unique equilibrium in nontrivial mixed strategies. The reason for looking for so many periods of play is to observe intermediate-term as well as short-term behavior. The data sets assembled report repeated play of 11 games, under a variety of experimental conditions, from the experiments of Malcolm and Lieberman (1965), Ochs (1995), O'Neill (1987), Rapoport and Boebel (1992), and Suppes and Atkinson (1960). In addition, a replication study of one of the conditions reported in Suppes and Atkinson (1960) was added. Thus a total of 12 data sets were analyzed.

Derivation of Predictions. To derive the models' predictions for these experiments computer simulations were conducted, designed to replicate the characteristics of each of the experimental settings. In each case the simulated players participated in the same number of rounds as the experimental subjects. Two hundred simulations were run in each game under different sets of parameters. At each round of each simulation the following steps were taken:

1. Simulated players were matched (using the matching procedure of the experiment being simulated).
2. The simulated players' strategies were randomly determined via Eq. 4.
3. Payoffs were determined using the payoff rule employed in the experiment in question.
4. Propensities were updated according to Eq. 5.

Parameter Estimation. A grid search with a mean squared deviation (MSD) criterion was conducted to estimate the value of the free parameters. That is, the simulations were run for a wide set of parameters, and the parameters that minimized the distance between the model and the data (minimized the model's MSD score) were selected, for each of the tests presented.

Main Results

Figure 4.1 presents the experimental and the simulation results in a sample of 3 of the 12 games. The payoff matrices are presented at the left-hand side of the

Data Equilibrium Reinforcement Learning

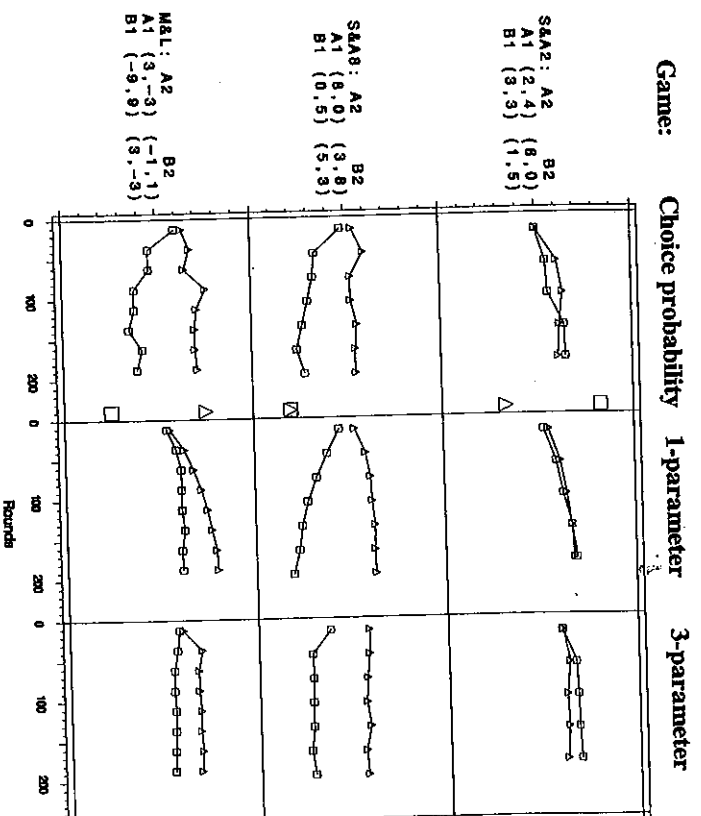


Fig. 4.1. Observed and predicted choice probabilities as a function of time over player type (P(A1) is represented by Δ and P(A2) by \square) in three of the games studied in Erev and Roth (1998). In the first two games (adapted from Suppes & Atkinson, 1960) each payoff unit increases winning probabilities (by 1/6 in Game S&A2, and by 1/8 in S&A8). In Game M&L (adapted from Malcolm & Lieberman, 1965) payoffs were directly converted to money.

figures. Each cell within the figure's frame is a graph that has the probability of a certain choice (ranging from 0 to 1) on the Y axis, and the rounds of the experiment (organized into blocks as in the data of that experiment) on the X axis. The first column of each figure summarizes the relevant experimental results. In each cell of this column, the mean probability with which Players 1 (row players) and Players 2 choose their first strategy (A) is plotted over time. (Player 1 choices are indicated by triangles, Player 2 choices by squares.) The

equilibrium predictions for Players 1 and 2, respectively, are given by the triangle and square at the far right of each cell in column 1.

The remaining two columns present the behavior of virtual subjects that behave according to the one-parameter and three-parameter models. The parameters utilized to derive these curves are the ones that best fit the data over the 12 games. The estimated parameters were $s(1) = 54$ for the one-parameter model, and $s(1) = 9$, $\epsilon = .2$, and $\phi = .1$ for the three-parameter model.

The top panel of Fig. 4.1 presents the results obtained in the mixed strategy experimental condition summarized in chap. 3 of Suppes and Atkinson (1960). The game played in this condition, referred to here as Game S&A2, has a unique mixed strategy equilibrium in which Player 1 chooses A1 with probability 1/3 and Player 2 chooses A2 with probability 5/6. It was played by 20 pairs of subjects for 200 rounds. The subjects were not informed that they were playing a two-person game. They were told that their task, in each of the 200 trials, was to predict which of two lights will be turned on. Subjects were run in pairs and the probability of a correct response was determined by the game payoff matrix (each point increases the probability of being correct by 1/6). Following each choice the subjects received accuracy feedback. Thus, although the subjects did not know that they were playing a game, the game is a description of the reinforcement structure (although subjects were not paid based on their performance, following Suppes and Atkinson's (1960) assumption that a "win" is a reinforcing event).

Suppes and Atkinson (1960) presented the choice proportions in blocks of 40 trials. The results (see top panel in Fig. 4.1) show that Player 2 appears to move toward the equilibrium prediction (the proportion of A2 choices increases with time). Player 1 initially moves away from the equilibrium. Only in the last two blocks is the proportion of A1 choices reduced.

The one-parameter model captures the initial learning trends (the simulated Player 2 moves immediately toward the equilibrium, whereas the simulated Player 1 starts moving in the opposite direction). Yet, this model fails to capture the late direction change in Player 2's behavior. The three-parameter model captures both the initial and the late trends.

The second panel in Fig. 4.1 summarizes the results obtained in a condition reported in chap. 4 of Suppes and Atkinson (1960), in which players knew that they were playing a game, but did not know the payoff matrix. At the equilibrium of this game (S&A8) both players choose A with probability .2. This game was played by 20 pairs of subjects for 210 rounds. Subjects were told that they were playing a two-person game in which they had to predict which of two lights will be turned on. They were told that the correct answer depends on their response, on the other subject's response, and on a random event. As in Game S&A2, the probability of correct responses were determined by the payoff matrix, which was not presented to the subjects.

The results (summarized by the proportions of A choices in blocks of 30 trials) are similar to the results obtained in Game S&A2. Whereas one of the

players (Player 2) quickly learns to approach the equilibrium prediction, the other (Player 1) initially moves away from the equilibrium. For this data also, the one-parameter model captures the initial trends whereas the three-parameter model captures the whole learning curve.

The third panel in Fig. 4.1 summarizes the results for the study conducted by Malcolm and Lieberman (1965), which was designed to test the descriptive power of the minimax (maximin) equilibrium prediction. The payoff matrix was explained to the subjects, and the payoffs units were chips that were converted to money at the conclusion of the experiment. Nine pairs of subjects participated in 200 replications of the game. At the equilibrium of this game Player 1 chooses A1 with probability 3/4 and Player 2 chooses A2 with probability 1/4.

Malcolm and Lieberman present the choice proportions in blocks of 25 trials. Experience led both players toward the equilibrium prediction, but Player 1 appears to learn faster, reaching equilibrium by the fourth block, whereas Player 2 approaches equilibrium slowly. Similar trends are exhibited by the simulated subjects.

Summary Statistics and Ex Ante Prediction of Behavior

Ideally we would like to be able to predict behavior at every level of aggregation or disaggregation, for every game, for any length of play. Because the models we consider are computational, we can use them to simulate each experiment and predict the probability of each action over time. Erev and Roth (1998) compared the predictions of different learning models and of equilibrium by computing the MSD of the predicted and observed behavior over time, for each game, both for all subjects and for individual pairs.

For each model and each of the 12 experimental data sets we performed two tests of descriptive power and one test of predictive power, as follows. First, we found the best parameters for minimizing the MSD over all games, and computed the MSD for each game using these parameters. Then we found the best parameters for minimizing the MSD for each of the 12 games separately (i.e., by looking at a model that replaces each parameter of the original model with 12 distinct parameters, one for each game). Finally, we tested the predictive power of each model on each of the 12 games by estimating the model's parameters on the data from the other 11 games, using the model to predict behavior in the game of interest, and comparing the predicted path of behavior with the observed path. The predictions reported here are for the entire path of play of the game over all periods.

Table 4.1 gives the results of the aggregate data (i.e., pooling over all subject pairs for each game). For brevity we omit the other analyses here.

Each of the first 12 rows of the table represents one of the games (initials reflect the authors who first experimented on each game). The numbers are the MSD scores (lower is better), with predictions being compared to data in real time (i.e., round n of the data is compared to round n of the simulations). The

Table 4.1
MSD Scores (100 x Mean Squared Deviation) Between the Distinct Predictions and the Experimental Results by Game and Over Games.

| Game | Random | Equilibrium | One-Parameter | | Three-Parameter | | | |
|-------|--------|-------------|---------------|-------------------|-----------------|----------|-------------------|------------|
| | | | Best Fit | By Game (12 par.) | Predicted | Best Fit | By Game (36 par.) | Prediction |
| S&A8 | 1.08 | 6.92 | 0.16 | 0.07 | 0.16 | 0.38 | 0.05 | 0.67 |
| S&A2 | 2.04 | 7.18 | 0.30 | 0.24 | 0.30 | 0.18 | 0.10 | 0.26 |
| S&A3u | 2.53 | 7.27 | 0.31 | 0.14 | 0.31 | 0.12 | 0.04 | 0.19 |
| S&A3k | 1.46 | 7.56 | 0.11 | 0.10 | 0.11 | 0.07 | 0.05 | 0.09 |
| S&A3n | 2.11 | 6.14 | 0.57 | 0.41 | 0.57 | 0.31 | 0.25 | 0.39 |
| M&L | 2.46 | 2.11 | 2.27 | 1.89 | 2.27 | 1.24 | 0.21 | 1.24 |
| On | 2.19 | 0.14 | 1.81 | 0.33 | 1.81 | 0.72 | 0.32 | 0.87 |
| R&B15 | 1.07 | 0.45 | 0.98 | 0.50 | 0.98 | 0.65 | 0.35 | 0.83 |
| R&B10 | 1.38 | 1.03 | 0.73 | 0.16 | 0.86 | 0.33 | 0.11 | 0.48 |
| O-9 | 3.88 | 2.22 | 2.71 | 2.34 | 2.71 | 1.54 | 1.34 | 1.54 |
| O-4 | 1.78 | 1.37 | 1.54 | 1.54 | 1.64 | 1.09 | 0.99 | 1.17 |
| O-1 | 0.45 | 0.45 | 0.48 | 0.41 | 0.48 | 0.48 | 0.37 | 0.51 |
| Mean | 1.87 | 3.57 | 1.00 | 0.68 | 1.02 | 0.59 | 0.35 | 0.69 |

table reports the results for random choice, for the equilibrium predictions, and for the two models presented earlier. For each model, the table reports the three comparisons in order; first the MSDs for the parameters that minimize the MSD averaged over all games, then for the parameters ($\times 12$) that separately minimize the MSDs for each of the 12 games, and then the MSD of the prediction for each game, using the parameters that best fit the other 11 games. The final row, which gives the mean over all games, gives a quick summary statistic by which the models can roughly be compared.

The table shows that the one-parameter reinforcement learning model outperforms the equilibrium prediction (and this remains true for *all* values of its one parameter; as $s(1)$ goes to infinity, the model approaches the random choice model). The model's descriptive and predictive power is further significantly improved by incorporating experimentation and forgetting into the three-parameter reinforcement model; this makes the model more responsive to a changing environment (i.e., an adaptive opponent).

Overall, the results support the notion that it may be possible to find learning models that can be usefully applied to a variety of games, rather than having to construct or estimate models separately for each game. For example, Table 4.1 shows that the three-parameter reinforcement model fit simultaneously to all games has a lower mean deviation (0.59) than does the one-parameter ($\times 12$)

model fitted to each game separately.

Initial Propensities, Adjustable Reference Point, and Generalizability

Whereas the three-parameter model provides a good prediction of behavior in the games considered in Erev and Roth (1998), it is clear that this model is oversimplified. In Roth and Erev (1995; also Slonim & Roth, 1998) we noted that the simplification assumption of uniform initial propensities has to be replaced with estimated initials in order to account for behavior in sequential bargaining games in which players in different countries have different initial propensities to use specific strategies.

In addition, it is easy to think about situations in which the assumed reinforcement function in the one- and three-parameter models ($R(x) = x - x_{\min}$) cannot provide a good approximation of behavior. For example, the addition of a dominated strategy that leads to a loss of \$100 with certainty to each of the games considered earlier is not likely to affect human behavior, but has a strong effect on the predictions of the two models (when x_{\min} is very small relative to all other payoffs these models predicts very slow learning). To address this limitation we proposed (in Erev & Roth, 1996a) a generalization of the three-parameter model in which payoffs are evaluated relative to an adjustable reference point. The generalized model has five parameters.

This generalization has little effect in the experiments considered by Erev and Roth (1998), but improves the model's predictions in other settings. The value of the reinforcement learning model with adjustable reference was demonstrated in the context of simplified poker games (Rapoport et al., 1997), market entry games and step-level public good games (Erev & Rapoport, 1998), and probability learning tasks (Bereby-Meyer & Erev, 1998).

To evaluate the generalizability and predictive power of the model with the parameters that best describe previous experiments, Roth, Slonim, Erev, and Bereby-Meyer (1997) studied 40 randomly selected 2×2 probabilistic constant-sum games under distinct information conditions. The payoff matrix in these games is defined by four probabilities (the probability that Player 1 wins in each cell). The random games were created by a uniform and independent random drawing of each of the four probabilities. Three pairs of subjects were run for 500 trials in each randomly selected game.

The main result of this extensive study is a demonstration that the *ex ante* prediction of simple reinforcement learning models of the behavior of the average pair is more accurate (has a smaller MSD score) than a prediction that is based on the average of the other two pairs that played the same game. Thus, it has "human power" of more than 2. In comparison, the human power of the equilibrium prediction is less than 1.

Table 4.2
A Matching Pennies Game

| Row Choice | Column Choice | |
|------------|---------------|------|
| | A | B |
| A | 1,-1 | -1,1 |
| B | -1,1 | 1,-1 |

VIOLATIONS OF SIMPLE REINFORCEMENT LEARNING MODELS

Despite the success that we have had in using reinforcement learning models to predict behavior in very simple strategic environments, it is clear that the model will need to be enriched if we are to extend it to more complex environments. Indeed, there is relatively large agreement among cognitive psychologists that reinforcement learning models are useful in predicting simple behavior, but fail to capture complex behavior. Six robust violations of reinforcement learning models are often used to justify this conclusion. These violations that have been observed both in psychological research and in experimental games are summarized here.

Overalternation and the "Gambler's Fallacy"

While studying the behavior of rats in a simple T-maze task, Tolman (1925) discovered a robust violation of the law of effect. In each trial of his study, rats had to choose an arm in the T maze. Whereas the law of effect predicts an increase in choice probability following a reward, Tolman found a decrease. His subjects tended to alternate even after winning a reward in one of the two sides. Interestingly, Rapoport and Budescu (1992) have found a similar phenomena in human behavior. In one of their experimental conditions, human subjects played a symmetrical zero-sum matching pennies game (cf. Table 4.2). Whereas their aggregate results (about equal choice of each alternative) are consistent with equilibrium and reinforcement learning, analysis of sequential dependencies reveal a clear violation of these two models.

Like Tolman's rats, Rapoport and Budescu's subjects exhibit a strong overalternation effect. In violation of the reinforcement learning prediction of a weak win-stay, lose-change dependency, their subjects tended to alternate even after winning.

Another robust sequential dependency that violates the law of effect was

Table 4.3
A PD Game

| Row choice | Column Choice | |
|------------|---------------|---------|
| | C | D |
| C | 1,1 | -10, 10 |
| D | 10,-10 | -1,-1 |

Note. Adapted from Rapoport and Channan (1965).

documented in probability learning studies. In a typical study (see the review in Lee, 1971) decision makers are asked to predict which of two events (L or H) will occur. In each trial only one event occurs, and the occurrence probabilities are fixed throughout the experiment. These studies reveal slow learning that can be approximated by Roth and Erev's model, and a violation of this model known as the "gambler's fallacy": At least in the beginning of some of the studies decision makers tend to predict a change in the state of the world.

Reciprocation

Rapoport and Channan (1965) studied behavior in seven repeated prisoner dilemma games (PDG). Table 4.3 presents one of the games they studied. In keeping with the definition of the PDG, each player has a dominant strategy (to choose D), but the "rational" outcome (the DD cell) is inefficient (both players could benefit from a move to the CC cell).

Rapoport and Channan's (1965) subjects were randomly matched at the beginning of the experiment and ran for 300 trials. Simple reinforcement learning models (like the Roth & Erev model) with only the stage game strategies "C" and "D" modeled predict convergence to the dominant strategy (see Bornstein et al., 1994). In violation of this prediction, an increase in C choices (cooperation) with time, and strong sequential dependencies, were observed in four of the seven games.²

Transfer

People's ability to transfer knowledge from one set of situations to another set is probably the toughest challenge for students of learning. Simple reinforcement learning models ignore transfer altogether. Roth and Erev's model, for example, was designed to address situations in which the learners play one game repeatedly. When decision makers participate in a few similar games during the same experimental block the model fails to predict the observed between-games facilitation.

For example, in Rapoport et al. (1998) subjects participated in 10 market entry games in each of 10 experimental blocks. The results reveal that much of the learning occurred before the second block. A single experience with each of the games was enough. This observation is inconsistent with the much slower prediction of a simple reinforcement learning model to this task.

A clearer example was documented in a binary categorization task research. In Kubovy and Healy (1977) and Kubovy, Rapoport and Tversky (1971) subjects were asked to categorize stimuli to one of two categories. Whereas each stimuli was presented only once, subjects learned and improved with practice.

Expectation-Driven Behavior

Whereas most comparisons suggest that reinforcement learning models provide a better approximation of behavior than expectation based models (see e.g., Camerer & Ho, chap. 3, this volume; Erev & Roth, 1998; Mookherjee & Sopher, 1997; Tang, 1996), it is clear that expectations affect behavior. For example, examination of large group multiple strategies games (Camerer & Ho, chap. 3, this volume; Cooper & Feltovich, 1996; Roth, Prasnikar, Okuno-Fujiwara, & Zamir, 1991; Van Huyck, Battalio, & Beil, 1991) reveals relatively fast learning that cannot be accounted for by reinforcement learning models (at least not given our quantification). Camerer and Ho (chap. 3, this volume) demonstrate that in the game they analyzed behavior is sensitive to outcomes of strategies that the decision maker has not selected. This sensitivity can be described as an attempt to maximize expected reward.

Imitation

Another reasonable explanation for the quick learning just discussed is that subjects imitate successful others. Bandura (1969) argued that vicarious learning (imitation) is inconsistent with reinforcement learning, and should be thought of as an independent learning process. Bandura supported this assertion with the observation that children tend to imitate others even when this behavior is not reinforced.

Direction Learning

Examination of the choice sequences of individual players often reveals consistent patterns that violate probabilistic models (like all reinforcement learning models). For instance, in games with a linearly ordered strategy space, some subjects consistently adjust their strategy in line to a direction-learning rule. Selten (1996) cited 10 studies in which some of the subjects can be categorized as "directional learners." To demonstrate the appeal of direction learning Selten presents an "archer" thought experiment. Obviously, a good archer has to adjust his or her behavior in line with a directional learning rule:

After missing the target to the left, the archer adjust the aim to the right. As noted by Selten, this trivial observation is not predicted by reinforcement learning models.

In line with Selten's suggestion, a strong knowledge of results (KR) effect was observed in motor learning research (Trowbridge & Cason, 1932). Subjects appear to learn faster when the feedback they receive includes directional and quantitative information in addition to the reinforcement.

THE COGNITIVE GAME THEORETIC EXPLANATION

According to the cognitive game-theoretic approach (Erev & Roth, 1998), it is convenient to decompose models of dynamic choice behavior into three submodels: (a) an abstraction of the cognitive strategies; (b) the incentive structure (the game); and (c) the learning/adaptation rule. Under this decomposition the apparent inconsistency between the descriptive power and the violations of reinforcement learning models can be explained by the distinction between the assumed cognitive strategies and the assumed learning rule. Cognitive strategies can be task-specific (like stage game strategies), or general adaptive rules (repeated game strategies like the tit-for-tat (TFT) strategy). The results summarized earlier are consistent with the hypothesis that the general learning rule can be approximated by a reinforcement learning model. The six violations previously summarized are then explained by the conjecture that they reflect a utilization of repeated game-cognitive strategies. That is, if reinforcement learning goes on among a rich set of strategies (and not merely among stage game strategies), then the phenomena we have been discussing no longer appear to violate the hypothesis that the learning process can be approximated as reinforcement learning.

It is important to emphasize that the cognitive game-theoretic approach is not suggested here as a testable model. Rather, like traditional game theory it is a theoretical framework that can be utilized to construct testable models. Our main assertion is that if cognitive strategies and games are assumed to be situation-specific (a common assumption in cognitive psychology and game theory), there is no need to assume a situation-specific learning model. And to the extent that cognitive strategies and the game can be assessed, the assumption of a general learning rule can be utilized to facilitate construction of useful models of choice behavior.

For the current discussion, the main difference between this explanation and alternative explanations that assume situation-specific learning processes is the assumed predictability of violations of simple reinforcement learning (i.e., reinforcement learning over actions rather than strategies). The cognitive game-theoretic explanation implies that the frequency of violations is predictable: Subjects are expected to learn to use strategies that lead to violations when they are positively reinforced and to stop using them when they impair

reinforcement. The situation-specific explanations can describe dynamic trends of this type (by fitting parameters), but currently do not have clear parameter-free predictions.

To evaluate the potential of the cognitive game-theoretic explanation this section discusses each of the observed violations. The discussion starts with a description of a cognitive strategy that could lead to the observed results. We then review studies that examine whether the magnitude of the violation (utilization of the assumed strategies) is affected by reinforcements. It is shown that in most cases clear learning effects that can be described as reinforcement learning among cognitive strategies are observed. In fact, the cognitive strategies explanation often coincides with the explanation provided (under different names) and supported in the psychological literature.

Sequential Strategies and Dependencies

Sequential dependencies can be a result of repeated game strategies. For example, rats (and humans) can follow an alternation strategy. Thirty years of alternation research is summarized by Dember and Fowler (1958). This research concludes that animals have a natural tendency to alternate. This tendency is assumed to be effective in most natural settings. (A rat that has eaten all the food in one location may profit from searching elsewhere rather than returning immediately.) Yet, when animals are put in a situation in which this strategy is inefficient, they can learn to avoid it. Moreover, when animals are put in a situation in which alternation is adaptive (as in Green, Price, & Hamburger's 1995 study in which pigeons played a chicken game against an alternating computer), they can learn to increase the probability of alternation.

Similar conclusions were reached in the human probability learning literature. It was found that the gambler's fallacy phenomena that impair earnings disappears as subjects gain experience (Estes, 1964).

Thus, it appears that rats, pigeons, and humans have an initial tendency to follow sequential strategies and learn to adjust the probability of using these strategies based on their effectiveness.

Note that this conclusion implies that at least initially players use repeated game strategies even in zero-sum games. Thus, it apparently contradicts the results summarized earlier that show that behavior in these games can be predicted by a reinforcement-learning model among the stage game strategies. To understand this contradiction Rapoport et al. (1997) derived the predictions of the Roth and Erev model for zero-sum games under two models of the strategy space: only stage game strategies or stage game plus alternation and a gambler's fallacy strategies. Their results indicate that the addition of the sequential strategies has very little effect on the aggregate predictions for such games. Thus, the paradox is resolved by the conclusion that in zero-sum games the approximation of the strategy space by the stage game strategies is inaccurate but robust.

Reciprocation Strategies

Reciprocation in repeated games is also observed widely enough for it to be sensible to think of reciprocation as a common repeated game strategy. For example, Axelrod (1984) argued that people follow the TFT strategy in PDG. Interestingly, Anatol Rapoport, who proposed the TFT strategy in Axelrod's tournament, did not use it to explain his experimental data. In fact, Rapoport and Chammah's (1965) data are not entirely consistent with the argument that players follow the TFT strategy. If players were simple TFT followers, reciprocation should have been observed immediately in all PDG. Rapoport and Chammah (1965) found a slow increase in reciprocation in some games and no increase in other games.

This observation is consistent with the view that TFT is a cognitive strategy. Erev and Roth (1996a) derived the prediction of a cognitive game-theoretic model for Rapoport and Chammah's (1965) games under the assumption of reinforcement learning among three cognitive strategies (TFT and the stage game strategies). This model reproduces the experimental results: it predicts an increase in reciprocation only when the experimental players learned to reciprocate.

Transfer: The Example of Cutoff Strategies

Transfer is predicted by the cognitive game-theoretic approach (and by all other cognitive models) because cognitive strategies can apply to many situations. For example, learning to follow a reciprocation strategy in one setting is expected to increase the probability of using this strategy in a different but similar setting. Whereas many questions regarding transfer are still open (like the measurement of similarity), it is clear that transfer is consistent with models that assume cognitive strategies. Moreover, in specific settings in which similarity can be approximate, cognitive game-theoretic models appear to provide a good prediction of learning.

For example, Rapoport et al. (1998) found that transfer between similar market entry games is predicted by a model that assumes reinforcement learning among cutoff strategies. A similar cutoff reinforcement learning model was found to reproduce all the robust regularities observed in binary categorization under uncertainty studies (Erev, 1998). In a typical binary categorization task a decision maker (DM) is presented with an observation (e.g., a height of an individual) that can come from one of two overlapping populations (e.g., heights of males and females). The DM's task is to classify the observation (e.g., decide if it is a male or a female) given a well-defined reward structure.

Previous research demonstrate that behavior in categorization tasks can be approximated by signal detection theory (SDT, Green & Swets, 1966). The relatively good approximation made SDT one of the most important tools for

applied psychologists (see e.g., Swensen, Hessel, & Herman, 1977; Wallsten & Gonzalez-Vallejo, 1994). This theory assumes that DMs consider cutoff strategies (e.g., respond male if the likelihood ratio of male exceeds a certain cutoff), and select the cutoff that maximize expected utility. Yet, careful examination of the difference between the predictions of SDT and the observed behavior reveals robust violations of this theory. Although the violations appear to decrease as DMs gain experience, they do not disappear even after thousands of decision trials with immediate feedback.

Different learning models were proposed to address the different violations of SDT (see Busemeyer & Myung, 1992 and an early review in Kubovy & Healy, 1977). Erev (1998) reviewed the known violations and compared previous explanations to a model that assumes reinforcement learning among cognitive cutoff strategies. This model is identical to the model presented earlier with the exception that at each period the DM selects a cutoff (that implies a decision) rather than one of the stage game strategies. Even without fitting parameters to specific experiments, the cutoff reinforcement learning model was found to capture 16 behavioral regularities and to outperform all previous explanations (typically post hoc alternative models).

Best Reply to Observed Statistics

The findings that people are affected by observed nonreinforcement events can be explained by the assumption that this information is utilized by certain cognitive strategies. For example, subjects who use a fictitious play strategy (Robinson, 1951) might explicitly calculate expected values (under the typically false assumption of a stationary world) and select the alternative that maximizes expected reward.

To test the hypothesis that best reply rules are strategies (rather than general learning models), it is useful to consider the evidence for best-reply behavior in games in which best reply to observed statistics is reinforcing. For example, in the Van Huyck et al. (1991) average opinion games, the optimal response is to match the population average. In these and similar games behavior can be described by a fictitious learning rule (Crawford, 1994), and as noted by Camerer and Ho players weigh nonreinforcement events. On the other hand, in market entry games in which matching other players is typically not reinforcing (and the best reply rule can lead to large losses), behavior is inconsistent with best-reply rules (Erev & Rapoport, 1998). An interesting and related analysis of the best-reply strategies is provided by Stahl (1996).

Imitation Strategies

Miller and Dollard (1941) provided an influential account of imitation within the reinforcement learning (operant conditioning) paradigm. In controlled experiments they established that the probability of imitation is a function of the

probability of reinforcement of previous imitative behavior. Bandura (1969) criticized this account and argued that a cognitive element has to be added to explain why children initially tend to imitate successful (and significant) others. He proposed a theory of imitation to account for this observation. Clearly, however, the required cognitive element can also be added by the assumption that people tend to follow cognitive imitation strategies that specify who should be imitated. This cognitive game-theoretic assumption is consistent with both Miller and Dollard's (1941) and Bandura's (1969) results. It is also consistent with the current view of learning by observation (Mazur, 1994).

Directional and Correctional Strategies

The observation that directional feedback and other information about outcomes facilitates learning can be accounted for by the assumption that people can follow directional strategies. For example, an archer can adjust to the observation that the wind carries arrows to the left by following a directional strategy that implies a correction to the right.

There are two types of evidence that suggest that directional learning is better abstracted as a cognitive strategy than as a learning rule. First, it turns out that the provision of feedback about outcomes after each trial during the learning period can have a negative effect (see e.g., Winstein & Schmidt, 1990). It appears that when this information is always provided subjects may learn to rely on it and, as a result, do not develop alternative and more robust strategies.

A second line of evidence includes observed behavior in games in which direction learning loses its effectiveness. Duffy and Nagel (1997; and see Nagel, chap. 6, this volume) found that the evidence for direction-learning behavior decreases with experience in a beauty contest (guessing) game when these strategies are not reinforced.

ON THE VALUE AND LIMITATIONS OF THE COGNITIVE GAME-THEORETIC APPROACH

As noted earlier, the fact that violations of simple reinforcement learning models can be accounted for within the cognitive game-theoretic framework does not imply that this framework is accurate. Moreover, the general framework is not testable; thus, the accuracy question is not relevant. The important question is "Is it useful?"

The results summarized earlier provide reasons to conjecture that this approach is likely to be useful. The most encouraging observation is the apparent robustness of the predicted choice probabilities to certain inaccuracies in the exact modeling of the cognitive strategies. It seems that even a rough approximation of these unobserved strategies can lead to nontrivial predictions. As noted in Erev and Roth (1998) the fact that learning curves in matrix games

in which players cannot reciprocate can be predicted from the assumption of stage game strategies does not imply that this assumption is an accurate approximation of the cognitive strategies. In fact, as the sequential strategies observed in these games suggest, this assumption is violated. The model appears to succeed because the predicted curves in this wide set of games are robust to the assumed cognitive strategies. To further explore this assertion we ran simulations of the constant-sum games studied in Roth et al. (1997) with simulated players who consider all the cognitive strategies mentioned that can be used in these games (alternation, gambler's fallacy, imitation and best reply to expectations). In line with Rapoport et al.'s (1997) results, addition of these strategies had a clear effect on the sequential dependencies, but only a mild effect on the aggregate choice probabilities. This does not imply that the exact cognitive strategies can be ignored. As noted earlier, in games that allow players to reciprocate an abstraction of a reciprocation strategy is needed to predict aggregate choice probabilities.

Another indication of the potential robustness of cognitive game-theoretic model comes from the study of models that assume learning among cutoff strategies. For example, Erev (1998) assumes that in binary categorization decisions DMs learn among 101 cutoff strategies. Whereas this assumption cannot be correct, Erev's model was found to account for 16 robust behavioral regularities that have been explained by six distinct models. Moreover, the cognitive game-theoretic model outperforms each of these six situation-specific models.

Finally, note that each of the cognitive strategies considered earlier was introduced in previous research. Thus, whereas some of these strategies are suggested here in a post hoc manner, it seems that it is possible to base a cognitive game-theoretic model on robust models developed in traditional cognitive research.

These results imply that even if there are infinitely many cognitive strategies that players might consider, a rough approximation of these strategies may be sufficient to predict choice probabilities. And this rough approximation can be obtained in an *ex ante* way for wide sets of well-defined situations.

CONCLUSIONS

The results just reviewed show that whereas simple reinforcement learning models (that assume learning among stage game strategies) provide a good approximation of behavior in some games, the predictions are violated in other games. The violations summarized earlier suggest that people (and other animals) "try to be smarter" than simple reinforcement learners. Each of the six violations of pure reinforcement learning can be summarized as an attempt to increase payoffs. Sequential dependencies, best reply, and directional learning can represent an attempt to efficiently reply to regularities in the world;

reciprocation maximizes earning by tacit cooperation; and transfer and imitation can facilitate efficient utilization of accumulated knowledge.

As noted in the beginning of the chapter, the contradiction between the predictive power and the clear violations of simple reinforcement learning models can be addressed by three distinct theoretical approaches: An "as if" model implies that players do not follow reinforcement learning rules but sometime behave as if they do. A game-specific learning approach implies that the extent to which players use reinforcement learning rules changes from game to game. And the cognitive game-theoretic approach implies that reinforcement learning determines choices among cognitive strategies. The results reviewed earlier summarize evidence that favors the cognitive game-theoretic account. This evidence suggests that the frequency of apparent violations of the law of effect may be predictable.

It should be emphasized that the support to the cognitive game-theoretic approach does not imply that the alternative approaches are wrong. Clearly, the reviewed evidence does not contradict the view that different models (or parameters) are needed to approximate learning in different games (Camerer & Ho, chap. 3, this volume; Selten, 1996). Rather, it suggests that reinforcement learning models can be used to predict which learning rule will best approximate behavior in a specific setting. For example, Camerer and Ho (chap. 3, this volume) suggest that it may be possible to describe the different models by a single functional form with game-specific parameters. In their research they show that in a simplified case the parameters can be estimated. The current results suggest that it may be possible to predict the value of the "best" parameters.

Strictly speaking, the current results do not even rule out the possibility that people never follow reinforcement learning rules. Roth and Erev's model is suggested here as an approximation of a slow adjustment process among cognitive strategies that appears to characterize behavior in games. Other approximations are, of course, possible. Yet, given the good fit provided by the current model to data described earlier, we conjecture that alternative models may use other theoretical terms (like beliefs) but are unlikely to be described by very distinct functional forms. (In Erev and Roth, 1998, we show that the main difference between our quantification of the law of effect and probabilistic fictitious play models can be summarized by three noncentral parameters.)

Finally, it is instructive to note that the cognitive game-theoretic approach represents a relatively minor modification of traditional game theory. The distinction between strategies and actions that we make here in connection with the reinforcement learning rule is a familiar one in traditional game theory. What we are proposing is that it may be possible to replace a general model of perfectly rational behavior (utility maximization, supplemented by assumptions of equilibrium behavior) with a general model of boundedly rational learning, supplemented by an empirically informed model of what strategies players consider.

The open question is how the cognitive strategies should be modeled. Advances in cognitive psychology and judgment and decision-making research suggest that good approximations can be obtained. This research suggests that in a wide set of situations, people tend to utilize relatively small sets of adaptive cognitive strategies. These strategies, often referred to as heuristics (Tversky & Kahneman, 1974), decision rules (Busemeyer & Myung, 1992; Payne, Betman, & Johnson, 1993), production rules (Anderson, 1982), or algorithms (Gigerenzer, 1996), are assumed to be used because they typically achieve good outcomes with relatively little cognitive effort. Yet, as demonstrated by Tversky and Kahneman (1974), in certain settings they can lead to counterproductive behavior (biases); in these settings people initially tend to follow inefficient strategies. Cognitive game theory may allow us to predict in which strategic environment initial behavior is likely to persist.

In summary, like traditional game theory, cognitive game theory assumes a general decision rule among situation-specific strategies. The major difference is the attempt to replace the rationally based assumption with a simple general learning rule, supplemented by an empirically based model of the strategy sets perceived by players. Our reading of the psychological literature suggests that the law of effect is the only property of learning sufficiently general so as to be able to drive learning in games with a wide variety of information conditions,³ and that initial strategies can be approximated. The fact that initial cognitive strategies are not always adaptive suggests that learning among these cognitive strategies can lead to nontrivial testable predictions.

ENDNOTES

1. In addition to these models we also considered a four-parameter variant of Camerer and Ho's model (see chap. 3, this volume). This model, which generalizes reinforcement learning and belief learning, did not outperform the three-parameter model on these data.
2. Similar results were obtained by Rapoport and Moskowitz (1966). And see Komorita and Parks (chap. 13, this volume) for a review of related studies.
3. Including situations in which individuals are not aware of the fact that they are learning to make more adaptive decisions. And repeated games are often played with very little awareness. For example, driving involves many repeated perceptual games: in a common scenario one of two drivers has to slow down to prevent an accident. Although we play this and similar games every day we are rarely aware of the fact that we make decisions and update our strategies. Yet, experimental studies of games of this type (Erev, Gopher, Itkin, & Greenshpan, 1995; Gilat, Meyer, Erev, & Gopher, 1997) show adaptive learning among cognitive strategies that can be described by Roth and Erev's quantification of the law of effect.

REFERENCES

- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 91, 112-149.
- Axelrod, R. (1984). *The evolution of cooperation*. New York, NY: Basic Books.
- Bandura, A. (1969). *Principles of behavior modification*. New York, NY: Holt, Rinehart & Winston.
- Bereby-Meyer, Y., & Erev, I. (1998). On learning to become a successful loser: Comparison of alternative abstractions of learning in the loss domain. *Journal of Mathematical Psychology*.
- Blackburn, J. M. (1936). *Acquisition of skill: An analysis of learning curves*. (HRB Rep. No. 73).
- Bornstein, G., Erev, I., & Goren, H. (1994). Learning processes and reciprocity in intergroup conflicts. *Journal of Conflict Resolution*, 38, 690-707.
- Brown, I. S., Clark, F. R., & Stein, L. (1958). A new technique for studying special generalization with voluntary responses. *Journal of Experimental Psychology*, 55, 359-362.
- Busemeyer, J. R., & Myung, I. J. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General*, 21, 177-194.
- Bush, R., & Mosteller, F. (1955). *Stochastic models for learning*. New York, NY: Wiley.
- Camerer, C. (1990). Behavioral game theory. In R. Hogarth (Ed.) *Insights in decision making: A tribute to Hillel J. Einhorn*. Chicago: University of Chicago Press.
- Cheung, Y., & Friedman, D. (1996). *A comparison of learning and replicator dynamics using experimental data*. Mimeo, University of California, Santa Cruz.
- Cooper, D., & Feltoovich, N. (1996). *Reinforcement-based learning vs. Bayesian learning: comparison*. (Working paper 305), University of Pittsburgh.
- Crawford, V. P. (1994). Adaptive dynamics in coordination games. *Econometrica*, 63, 103-143.
- Dember, W. N., & Fowler, H. (1958). Spontaneous alternation behavior. *Psychological Bulletin*, 55, 412-428.
- Duffy, J., & Nagel, R. (1997). On the robustness of behavior in experimental "P-beauty contest" games. *Economic Journal*, 107, 1684-1700.
- Erev, I. (1998). Signal detection by human observers: A cutoff reinforcement learning model of categorization decisions under uncertainty. *Psychological Review*, 105(2), 280-298.
- Erev, I., Gopher, D., Itkin, R., & Greenshpan, Y. (1995). Toward a generalization of signal detection theory to N-person games: The example of two-person safety problem. *Journal of Mathematical Psychology*, 39, 360-375.
- Erev, I., & Rapoport, A. (1998). Coordination, "magic," and reinforcement learning in a market entry game. *Games and Economic Behavior*, 23, 146-175.
- Erev, I., & Roth, A. (1996a). *A cognitive game theoretic analysis of reciprocity*. Paper presented at the workshop on Games and Human Behavior in the honor of Amnon's Rapoport 60th birthday, Chapel Hill, NC.
- Erev, I., & Roth, A. (1996b). *On the need for low rationality, cognitive game theory: Reinforcement learning in experimental games with unique, mixed strategy equilibria*. Mimeo, The University of Pittsburgh.
- Erev, I., & Roth, A. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic*

Review.

- Estes, W. K. (1964). Probability learning. In A. W. Melton (Ed.), *Categories of Human Learning* (pp. 89-128). New York, NY: Academic Press.
- Fudenberg, D., & Levine, D. K. (1996). *Theory of learning in games* [draft]. Available at: <http://levine.ssrnet.uchicago.edu/papers/contents.htm>.
- Gigerenzer, G. (1996). *Introducing satisfying models of inference and how they affect our notions of sound reasoning and rationality*. Paper presented at the meeting of the society of Judgment and Decision Making, Chicago, IL.
- Gilat, S., Meyer, J., Erev, I., & Gopher, D. (1997). Beyond Bayes' theorem: The effect of base rate information in consensus games. *Journal of Experimental Psychology: Applied*, 2, 83-104.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley. (Reprinted in 1988; Los Altos, CA: Peninsula Publishers)
- Green, L., Price, P. C., & Hamburger, M. (1995). Prisoner dilemma and the pigeon: Control by immediate consequences. *Journal of Experimental Analysis of Behavior*, 64, 1-17.
- Harley, C. B. (1981). Learning the evolutionary stable strategy. *Journal of Theoretical Biology*, 89, 611-633.
- Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior*, 13, 243-266.
- Kubovy, M., & Healy, A. F. (1977). The decision rule in probabilistic categorization: What it is and how is it learned. *Journal of Experimental Psychology: General*, 106, 427-446.
- Kubovy, M., Rapoport, A., & Tversky, A. (1971). Deterministic vs. probabilistic strategies in detection. *Perception & Psychophysics*, 9, 427-429.
- Lee, W. (1971). *Decision theory and human behavior*. New York, NY: Wiley.
- Luce, D. R. (1959). *Individual choice behavior*. New York, NY: Wiley.
- Malcolm, D., & Lieberman, B. (1965). The behavior of responsive individuals playing a two-person, zero-sum game requiring the use of mixed strategies. *Psychonomic Science*, 373-374.
- Mazur, J. E. (1994). *Learning and behavior*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Miller, N. E., & Dollard, J. (1941). *Social learning and imitation*. New Haven, CT: Yale University Press.
- Moocklejee, D., & Sopher, B. (1997). Learning and decision costs in experimental constant sum games. *Games and Economic Behavior*, 19, 97-132.
- Ochs, J. (1995). Simple games with unique mixed strategy equilibrium: An experimental study. *Games and Economic Behavior*, 10, 202-217.
- O'Neill, B. (1987). Nonmetric test of the minimax theory of two-person zerosum games. *Proceedings of the National Academy of Sciences, USA*, 84, 2106-2109.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, England: Cambridge University Press.
- Rapoport, Am., & Boebel, R. B. (1992). Mixed strategies in strictly competitive games: A further test of the minimax hypothesis. *Games and Economic Behavior*, 4, 261-283.
- Rapoport, Am., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General*, 121, 352-363.
- Rapoport, Am., & Moskowitz, A. (1966). Experimental studies of stochastic models for the prisoner dilemma. *Behavioral Science*, 11, 444-458.
- Rapoport, Am., Erev, I., Abraham, E. V., & Olson, D. E. (1997). Randomization and

- adaptive learning in a simplified poker game. *Organizational Behavior and Human Decision Processes*, 69, 31-49.
- Rapoport, Am., Seale, D. A., Erev, I., & Sundali, J. A. (1998). Coordination success in market entry games: Tests of equilibrium and adaptive learning models. *Management Science*, 44, 119-141.
- Rapoport, Am., & Chammah, A. M. (1965). *Prisoner dilemma*. University of Michigan Press.
- Robinson, J. (1951). An iterative method of solving a game. *Annals of Mathematics*, 54, 296-301.
- Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* [Special Issue: Nobel Symposium], 8, 164-212.
- Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *American Economic Review*, 81, 1068-1095.
- Roth, A. E., Slonim, B., Erev, I., & Barely-Meyer, Y. (1997). *On the predictive power of reinforcement learning models: Predicting behavior in randomly selected games*. Mimeo, University of Pittsburgh.
- Selten (1996). *Learning direction theory*. Paper presented at the workshop on Games and Human Behavior in honor of Amnon's Rapoport 60th birthday, Chapel Hill, NC.
- Slonim, R., & Roth, A. E. (1996). *Financial incentives and learning in ultimatum and market games: An experiment in the Slovak republic*. *Econometrica*.
- Stahl, D. O. (1996). *Evidence based rule learning in symmetric normal-form games*. Working Paper, University of Texas.
- Suppes, P., & Atkinson, R. C. (1960). *Markov learning models for multiperson interactions*. Palo Alto, CA: Stanford University Press.
- Svensen, R. G., Hessel, S. J., & Herman, P. G. (1977). Omission in radiology: Faulty search or stringent reporting criteria? *Radiology*, 123, 563-567.
- Tang, F. (1996). *Anticipatory learning in two-person games: An experimental study*. Part II. *Learning*. Discussion Paper B-363, University of Bonn, Germany.
- Thorndike, E. L. (1898). *Animal intelligence: An experimental study of the associative processes in animals*. Psychological Monographs, 2.
- Tolman, E. C. (1925). Purpose and cognition: The determiners of animal learning. *Psychological Review*, 32, 285-297.
- Trowbridge, M. H., & Cason, H. (1932). An experimental test of Thorndike's theory of learning. *Journal of General Psychology*, 7, 245-260.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Van Huyck, J., Battalio, R., & Beil, R. (1991). Strategic uncertainty, equilibrium selection, and coordination failure in average opinion. *Quarterly Journal of Economics*, 106, 885-909.
- Wallsten, T. S., & Gonzalez-Vallejo, C. (1994). Statement verification: A stochastic model of judgment and response. *Psychological Review*, 101, 490-504.
- Winstein, C. J., & Schmidt, R. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 677-691.