GAMES AND ECONOMIC BEHAVIOR **23,** 146–175 (1998)
ARTICLE NO. GA970619

Coordination, ''Magic,'' and Reinforcement Learning in a Market Entry Game*

Ido Erev†

Technion—Israel Institute of Technology, Haifa, 32000 Israel

and

Amnon Rapoport

University of Arizona, Tucson, Arizona 85721-0001

Received June 10, 1996

Previous experimental studies have documented quick convergence to equilibrium play in market entry games with a large number of agents. The present study examines the effect of the available information in a 12-player game in an attempt to account for these findings. In line with the prediction of a simple reinforcement learning model (Roth and Erev, 1995, *Games Econ. Behav.* **8**, 164–212), quick convergence to equilibrium is observed even given minimal information (unknown payoff rule). However, in violation of the basic model, information concerning other players' payoff increases the number of entrants. The information effect can be described by a variant of the basic reinforcement learning model assuming that the additional information changes the player's reference point. *Journal of Economic Literature* Classification Numbers: C7, C92.

Whereas many experimental studies of noncooperative game theory suggest that the Nash equilibrium solution does not provide an accurate description of human interactive decisions (e.g., Kagel and Roth, 1995; Camerer, 1990), several recent studies (Kahneman, 1988; Rapoport, 1995; Rapoport *et al.*, in press; Sundali *et al.*, 1995) have shown that interactive decisions in a class of market entry games with a relatively large number of

*This work was supported in part by Grant 604-94-1 on market entry games from the Israel Science Foundation, awarded to Ido Erev and Amnon Rapoport, and NSF Grant SBR-95-12724, awarded to Amnon Rapoport and Kenneth Koput. The authors wish to thank Tamar Landsbeger, Einat Duvdevani, and Shmuel Bar-Shai for help in data collection and constructive comments. Part of this research was conducted when Erev was on sabbatical at the Department of Economics, University of Pittsburgh.

† E-mail: erev+@pitt.edu.

0899-8256/98 $25.00

agents are accounted for surprisingly well by the equilibrium solution. To Kahneman, who seemed to be bewildered by the results, the behavioral regularities found in this game looked "like magic" (Kahneman, 1988, p. 12). Subsequent studies of the market entry game by Rapoport and his associates, which systematically manipulated the information structure of the game and the domain of payoffs (gains vs. losses), have shown that this "magic" is robust. Under a wide variety of experimental conditions, interacting players in large groups playing the market entry game with no communication rapidly achieve a degree of tacit coordination, which is accounted for on the aggregate level by the Nash equilibrium solution.

Dozens of studies have examined interactive behavior in situations that facilitate deviations from rational play (e.g., experimental studies of the Prisoner's Dilemma, the two-person ultimatum bargaining game, and the Centipede game). In contrast, considerably fewer experimental studies have examined the structural and contextual (Crawford, 1995) principles that induce interactive behavior to converge to equilibrium. The present study focuses on the regularities observed in the market entry game in an attempt to understand the process that leads aggregate decisions to equilibrium. In particular, it tests the general hypothesis that the results of several market entry games can be accounted for by an adaptive learning model recently proposed and extensively tested by Roth and Erev (Roth and Erev, 1995, and Erev and Roth, in press).

Rapoport *et al.* (in press) have shown that the Roth−Erev model can be extended to account for the market entry game's results. They have considered games in which the market size changes randomly from period to period, and they found that, with two additional assumptions, the Roth−Erev model can reproduce major features of their experimental results. The present study considers a different experimental design in which the market capacity is held constant across all periods. By manipulating the information structure in a between-subjects design, it provides a more stringent test of the Roth−Erev reinforcement learning model.

Several issues are addressed in the present research. We first examine whether a simple two-strategy reinforcement learning process (the Roth−Erev basic model without any modifications) can result in convergence to equilibrium in the market entry game. Computer simulations of the learning model indicate that the answer to this question is positive.

We then describe in detail a controlled experiment specifically designed to explore the main differences between the predictions of the reinforcement-based learning model and human behavior. The results of this experiment show that the basic two-strategy learning model provides a good approximation to the aggregate results. However, information about other players' payoff, that should have no effect according to the Roth−Erev model, does influence behavior in this experiment. Model comparisons suggest that this effect can be described by the assumption that information concerning other player's payoff affects the classification of the reinforcements as "positive" or "negative." The information effect might also be explained by an assumption that subjects use imitation and best reply rule, but the support for this view in the current data is weak.

The final section considers general implications and then applies the original and new variant of the adaptive learning model to the results of previous market entry studies and similar step-level public good games.

MARKET ENTRY GAMES

In the class of games under investigation, each player must decide privately and anonymously whether to enter a certain market. The player's payoff is the difference between two components: market profit and entry costs. The game is symmetric with respect to the market profits, but entry costs may differ from one player to another. Formally, using the terminology of Selten and Güth (1982), we consider a noncooperative n-person game in strategic form where each player i, $i = 1, 2, \ldots, n$, has two pure strategies, namely, stay out or enter. In addition, each player i has entry costs C_i. It is further assumed that each player who enters the market receives a payoff R_m, and each player who stays out receives a payoff S_m, where m $(0 \leq m \leq n)$ denotes the number of entrants.

Under these assumptions, the individual payoff function H_i of player i has the form

$$H_i = \begin{cases} R_m - C_i & \text{if player } i \text{ enters} \\ S_m & \text{if player stays out.} \end{cases}$$

The term $A_m = R_m - S_{m-1}$ is the incentive to enter. It is natural to assume that A_m is a nondecreasing function of m: $A_1 \geq A_2 \geq \cdots \geq A_n$. This assumption means that the larger the number of entrants, the smaller the incentive to enter the market.

As pointed out first by Selten and Güth and later by Gary-Bobo (1990), the market entry game can be interpreted as the "truncation" or first stage of a two-stage game of the following form. In the first stage of the game (entry stage) the n players simultaneously choose either to stay out or enter. The vector of n binary choices is then revealed to all n players. In the second stage (supply stage) all the players who entered the market in stage 1 simultaneously make some other decision (e.g., determine a supply level in a Cournot game). If all the subgames in stage 2 have uniquely defined equilibria, the market entry game is then obtained as a truncation of the original two-stage game.

The market entry game may be used to model asymmetric oligopolistic market situations in which participating firms must make a preliminary entry decision and then compete in the second stage. The linear Cournot oligopoly model studied theoretically by Selten and Güth is one example, as are other oligopoly models where prices, advertising expenditures, or quality parameters are the decision variables. Another class of applications includes two-stage processes where the first stage is an innovation rather than an entry stage (Selten and Güth, 1982). In the first stage, each of n suppliers can adopt a production method that saves labor but requires investment costs. These costs may vary from one supplier to another as a function of size, technical ability, and the R & D budget. The second-stage subgame equilibria will depend on the number m of suppliers who have adopted the new production technology in stage 1. Traffic problems constitute yet another class of potential applications. In many countries with heavily congested roads one faces a decision whether to take a train (stay out) or drive a car (enter) from one destination to another. If the train is spacious enough, the utility of choosing it is relatively fixed, whereas the utility of traveling by car decreases with the number of cars on the road (i.e., entrants).

Previous Research

Previous experimental studies of the market entry game examined a special case with symmetric players, zero entry costs, and an incentive to enter the market that decreases linearly in the number of entrants. In all of these studies, R_m assumed the form

$$R_m = 1 + 2(c - m),$$

where $c(0 < c \leq n)$ is a fixed and known constant interpreted as the "market capacity." Individual entry costs and payoff for staying out were set at constant values:

$$C_i = 0 \qquad \text{for all } i$$

and

$$S_m = v \qquad \text{for all } m.$$

In the two studies of Sundali *et al*. (1995) the (constant) payoff for staying out was set at $v = 1$. Rapoport *et al*. (in press) set v at some either relatively large positive or negative value. Thus, the payoff function in all of these studies assumed the form

$$H_i = \begin{cases} v & \text{if stay out} \\ 1 + 2(c - m) & \text{if enter.} \end{cases}$$

In all of these studies, communication before or during the game was prohibited. The stage game was iterated for T periods. A special characteristic of these previous studies, not shared by the present study, is that the market capacity value c was varied randomly from period to period. In particular, once some value of c was displayed on the computer terminal, each of the n subjects made her decision whether or not to enter privately and anonymously. When all the subjects entered their decisions, the computer terminal displayed the number of entrants (m), the individual payoff (H_i), and the accumulated payoff of subject i from the beginning of the experiment. Then a new stage game started with a different randomly chosen value of c. Altogether, 10 different values of c were displayed in a random order in each block of 10 periods. The same values of c were then iterated in a different random order on each of the subsequent blocks. Each group of subjects completed 10 blocks of 10 periods each for a total of $T = 100$ periods.

In Rapoport (1995) and Sundali *et al.* (1995) (and in the current study) the payoff for staying out is set at $v = 1$. If $v = 1$, then there are $n!/(c!(n-c)!)$ weak pure-strategy equilibria with c entrants and $n - c$ nonentrants, as well as $n!/((c-1)!(n-c+1)!)$ weak pure-strategy equilibria with $c - 1$ entrants and $n - c + 1$ nonentrants. In addition, there is a unique symmetric mixed-strategy equilibrium where, given the commonly known value of c, each player enters with probability $p(c)$ and stays out with probability $1 - p(c)$, where

$$p(c) = (c-1)/(n-1).$$

It is easy to verify that $c - 1 \leq np(c) \leq c$. Thus, whether pure or mixed-strategy equilibria are used, the mean number of entrants should be between c and $c - 1$.[1]

The major findings of the previous studies can be briefly summarized. Positive and highly significant correlations between the 10 pairs of c and m values were found on each block. For groups of $n = 20$ subjects, the correlations were around 0.90. When several different groups of subjects were combined ($n = 60$), the correlations increased to about 0.98. Rapid convergence to the equilibrium was already achieved on the first block. These highly regular and replicable results on the aggregate level were observed together with large within-subjects variability in the decisions across the 10 iterations of each c, and considerable between-subjects variability in decision profiles that did not decrease with experience. Systematic manipulation of the domain of payoffs, achieved by setting

[1]These equilibria are inefficient (as total payoff is maximized when $m = c/2$). Coordination market games (with Pareto efficient equilibria) were studied by Ochs (1990) and Meyer *et al.* (1992); see Ochs (1995b) for review.

$v = -10$ and $v = -6$ (domain of losses) or $v = 6$ (domain of gains), did not affect the results (Rapoport *et al.*, in press).

Although the values of m rapidly converged to c or $c - 1$ on the aggregate level (when $v = 1$), no support was found for either the pure-strategy or symmetrical mixed-strategy equilibria on the individual level. In violation of the pure-strategy equilibrium prediction that implies static decision policies, large within-subjects variability was observed. And in violation of the symmetrical mixed-strategy equilibrium prediction, the between-subjects standard deviations of number of entries for every value of c were always larger than $(p(c)(1 - p(c))n)^{1/2}$, the value predicted at this equilibrium. In addition, in 20–30% of all cases, the subject alternated his/her decision for the same value of c presented in successive blocks of trials. Taken together, these results suggest the operation of some adaptive learning process that, on the aggregate level, converges rapidly to the equilibrium solution, while still allowing for changes in individual decisions for the same market capacity presented several times as well as differences between individual decision profiles.

THE ROTH–EREV ADAPTIVE LEARNING MODEL

The adaptive learning model of Roth and Erev is built upon a quantification of Thorndike's (1898) law of effect (similar quantifications were suggested by Herrnstein, 1970, and Harley, 1981). The law of effect states that good outcomes (reinforcements) associated with playing a particular strategy increase the probability that this strategy will be chosen again. To the basic quantification, Roth and Erev (1995) and Erev and Roth (1996) added quantifications of three additional important characteristics of human and animal learning processes. For expository purposes, it is convenient to present the basic quantification first. It can be described by the following three assumptions:

A1. *Initial propensities*: At time $t = 1$ (before any experience has been acquired), each player i has an initial propensity to play his kth pure strategy, given by some real number $q_{ik}(1)$.

A2. *Reinforcements*: If player i plays his kth pure strategy at time t and receives a nonnegative payoff of x, then the propensity to play strategy k is updated by setting

$$q_{ik}(t + 1) = q_{ik}(t) + x,$$

whereas for all other pure strategies j,

$$q_{ij}(t + 1) = q_{ij}(t).$$

A3. *Probabilistic response rule*: The probability $p_{ik}(t)$ that player i plays his kth pure strategy at time t is given by

$$p_{ik}(t) = q_{ik}(t)/\Sigma q_{ij}(t),$$

where summation is taken over all of player i's pure strategies j. Thus, pure strategies which have been played and have met with success tend over time to be played with greater frequency than those which have met with failure.

The three additional characteristics of the learning process that were explicitly modeled by Erev and Roth (1996) are generalization (and experimentation), recency, and effect of reference points. These factors were quantified in the following generalized version of assumption A2:

A2*. If player i plays his kth pure strategy at time t and receives a payoff of x, then the propensity to play any strategy j is updated by setting

$$q_{ij}(t+1) = \max\left[v, (1-\phi)q_{ij}(t) + E_k(j, R_t(x))\right].$$

In assumption A2*, ϕ is a forgetting (or recency) parameter which slowly reduces the importance of past experience, R is a function which translates payoffs into rewards, and E is a function which determines how the experience of playing strategy k and receiving the reward $R(x)$ is generalized to update each strategy j. ($v > 0$ is a technical parameter to ensure that the propensities remain positive.)

Erev and Roth (1996) utilized empirical observations to determine the shape of the reinforcement and generalization functions. Following Herrnstein's (1961) demonstration of a linear relation between reinforcements and choice probabilities, they set

$$R_t(x) = x - \rho(t),$$

where $\rho(t)$ is the reference point on trial t. Erev and Roth (1996) denoted the reference point at the beginning of the experiment by $\rho(1)$ and assumed that the reference point is updated by the following linear weighing function:

$$\rho(t+1) = \begin{cases} (1-w^{+})\rho(t) + (w^{+})x & \text{if } x \geq \rho(t) \\ (1-w^{-})\rho(t) + (w^{-})x & \text{if } x < \rho(t), \end{cases}$$

where w^{+} and w^{-} are the weights assigned to positive and negative reinforcements, respectively.

For games with two pure strategies such as the present version of the model, the generalization function is reduced to

$$E_k(j, R(x)) = \begin{cases} R(x)(1-\varepsilon) & \text{if } j = k \\ R(x)\varepsilon & \text{otherwise,} \end{cases}$$

where $0 \le \varepsilon < 0.5$ is a generalization (and experimentation) parameter.

It is worth noticing that the Roth–Erev model makes no use of the notion of best reply, which many adaptive learning models incorporate. More important for our purpose, the model incorporates no other information about the behavior and payoffs of the other players or the "social history" of the game; as such, it is applicable in its present form to both individual and interactive decision-making tasks. We exploit this property in the variants of the market entry game studied below.

The Parameters of the Model

To reduce the number of parameters in their learning model, Erev and Roth (1996) noted that under assumption A3 the initial propensity parameters can be written as

$$q_{ik}(1) = p_{ik}(1)\left[\sum_{j=1}^{m} q_{ij}(t)\right].$$

Setting $S(1) = \sum_{j=1}^{m} q_j(t)/(X)$, where X is the game average outcome given random decisions, and $S(1)$ is defined as a "strength parameter," the task of determining the initial propensities is therefore reduced to the task of determining $S(1)$ and the initial choice probabilities. Uniform initial choice probabilities are assumed in the present study.

After reducing the number of initial propensities, the general model includes seven parameters[2] $(\varepsilon, \upsilon, \phi, S(1), \rho(1), w^+,$ and $w^-)$. Erev and Roth (1996) found that in the context of strategic games with unique mixed strategy equilibrium in which players cannot reciprocate, the model's descriptive power is relatively insensitive to the exact values of the parameters. They found high correspondence (correlations above 0.9 and mean squared deviations below 0.015) between predicted and observed learning functions, given a wide set of parameters.

The present paper avoids the free parameter problem altogether by utilizing the parameter set that minimized the distance between the model predictions and the data in the study of Erev and Roth (1996). The exact

[2] The reference point's parameters ($\rho(1)$, w^+; and w^-) and the technical parameter (υ) are needed to account for learning in the loss domain (and w^- has to exceed w^+ to insure that the power law of practice is satisfied); initial strength ($S(1)$) is needed to model initial tendencies; and experimentation (ε) and recency (ϕ) parameters are required to allow adaptation in a changing environment.

values taken from Table 3 of Erev and Roth (1996) are $S(1) = 3$, $\varepsilon = 0.2$, $\phi = 0.001$, $\rho(1) = 0$, $w^{+} = 0.01$, $w^{-} = 0.02$, and $v = 0.0001$. Whatever results we report below could only be improved by a judicious selection of new parameter values.

Despite the relatively large number of parameters, the Roth–Erev model is falisifiable. Erev and Roth (in press) noted that their basic model has to be extended to account for behavior in games in which players can reciprocate. For example, the basic model is violated by the observation that players learn to reciprocate by alternating in a repeated Chicken game (see Rapoport *et al*., 1976). Note that Chicken can be formulated as an example of a two-person market entry game. The present paper focuses on n-person games in which reciprocation is not possible.

THE EXPERIMENT

In all previous studies of the market entry game, the market capacity was randomly varied from period to period. This design feature resulted in a very complex market entry game in which the learning process is likely to be affected by generalization from one market size to another. For example, suppose that $c = 11$ and $m = 8$ on period t. Then if $c = 7$ on period $t + 1$, most players may attempt to exploit the departure from equilibrium on trial t by entering on $c = 7$, thereby raising the number of entrants above c. Other, possibly more farseeing players may anticipate this "overreaction" and decide to stay out. Although the same situation occurs if the same market capacity is presented on consecutive periods, the effect of trial-to-trial variation in c is eliminated. Another characteristic of the "random market size" design is related to the observability of the learning process. As noted by Rapoport *et al*. (in press), this design appears to lead the subjects to consider cutoff strategies, and even random selection among strategies of this type can lead to equilibrium-like behavior.

Therefore, to facilitate understanding of the adaptive learning process, the present study focuses on a considerably simpler version of the market entry game in which the market capacity is constant across iterations of the stage game. In particular, our subjects played two market entry games in a within-subjects design, one with $c = 4$ and the other with $c = 8$ ($n = 12$ in both games).

A second major departure from all previous studies was in our experimental manipulation of the game's information structure. Recall that in the two studies by Sundali *et al*. (1995, Exp. 2) and Rapoport *et al*. (in press), complete outcome feedback was provided at the end of each period about the number of entrants, individual payoff for the period, and accumulated individual payoff from the beginning of the experiment. In contrast, the information structure in the present study was systematically degraded by withholding information about the payoff rule determining individual payoffs and privatizing the individual payoff information. In information Condition 3, the payoff rule was explained to the subjects and individual payoffs were publicly displayed, as in the previous studies. In Condition 2, individual payoffs were publicly displayed, as in Condition 3, but no information about the payoff rule was given. Finally, in information Condition 1 the payoff rule was not explained, as in Condition 2, and information about individual payoff for the period was only privately given. These three information conditions were presented in order to test an important implication of the Roth−Erev adaptive learning model, which, as mentioned above, does not consider the payoff rule and the public nature of the payoff feedback, treating these three rather different conditions as equivalent.

METHOD

Subjects

The subjects were 144 Israeli undergraduate students at the Technion. They were recruited by campus advertisements promising monetary reward for participation in a group decision-making task. The subjects participated in the experiment in groups of $n = 12$; 4 groups participated in each of 3 information conditions for a total of 12 groups.

Procedure

Upon arrival at the laboratory, the subjects were seated in a single room and were handed written instructions.[3] The instructions explained that each participant would earn 20 IS (about $7) for showing up at the experiment and that he or she could either gain or lose more money during the experiment. The exact payoff was said to depend on the outcome of a (randomly determined) single round of play. Subjects were told that in each round they would have to make a binary decision (choose one of two states of an electrical switch).

The instructions for Condition 3 contained a detailed explanation of the payoff rule. A description of the payoff rule was not provided to the subjects in the other two conditions. In fact, subjects in these two latter conditions did not receive any information that they were participating in a noncooperative *n*-person game. (However, because they were seated to-

[3]Instructions to the subjects (translated from Hebrew) can be obtained from the first author upon request.

gether in the same room, the subjects might have deduced that their payoff was affected in some way by the decisions of others.)

Each group of subjects played 20 rounds (iterations) of each of the two 12-person market entry games (one with $c = 4$ and the other with $c = 8$). The presentation order of the two games was balanced across groups. The number of rounds was not made known to subjects. Each subject had an electrical switch controlling a light bulb on a board in front of the experimenter. The subjects could not see the board. At the beginning of each round subjects had 20 seconds to decide between their switch's two states (0 and 1). Whereas state "1" stood for entering the market, this rule was only known to the subjects in Condition 3.

An outcome feedback was provided as follows. In Condition 3, the experimenter announced the number of entrants and the exact payoffs. In Condition 2, the experimenter only announced the payoff for each of the two decisions. Finally, in Condition 1 each subject received private information about his or her payoff.

At the end of each round of play, the subjects were given 10 seconds to record their outcome. Following the last round, the round that determined the actual payoff was randomly determined. Subjects were then debriefed, paid, and dismissed individually from the laboratory.

RESULTS

We start by examining the implications of the Roth−Erev model for short-term play of the market entry game. Because of the probabilistic nature of the predicted response, which prohibits the derivation of closed-form expressions, we revert to simulations (see, e.g., Bush and Mosteller, 1955). Five hundred different simulations were conducted to generate mean learning functions. In each simulation, groups of 12 simulated players (called "stat-rats" by Bush and Mosteller) each played the market entry game with market capacity of either $c = 4$ or $c = 8$. Each game was played for only 20 rounds, mimicking the behavior of the real subjects.

On each round of the simulation play, the following steps were implemented:

1. The stat-rats' strategies were randomly determined in accordance with assumption A3.

2. Individual payoffs were calculated in accordance with the game payoff scheme.

3. Individual propensities were then updated, given the payoff for the round, in accordance with assumption A2*.

4. The reference point for the next round was determined.

Summary statistics of the simulation results are depicted in the left-hand column of Fig. 1. The top panel (Fig. 1a) displays the mean number of entrants. These results are shown in terms of four blocks of five rounds each. Rapid convergence to equilibrium is clearly observed for both games with $c = 4$ and $c = 8$.

Figure 1b portrays another summary statistic of the simulation results, namely, the mean absolute deviation between the number of entrants and the expected number of entrants under the symmetrical mixed-strategy equilibrium solution (3.27 when $c = 4$ and 7.64 when $c = 8$). For both games, the functions are seen to decrease with experience in accordance with the results shown at the top panel. The down trend implies that convergence to equilibrium occurs on the individual level, and it is not a function of aggregation over individual simulations.

Figure 1c shows the probability of alternation between rounds t and $t + 1 (t = 1, 2, \ldots, 19)$. Note that at the mixed-strategy equilibrium, the expected probability of alternation is $1 - (7/11)^2 + (4/11)^2 = 0.456$ for $c = 8$, and $1 - (3/11)^2 + (8/11)^2 = 0.397$ for $c = 4$. In contrast, the expected probability of alternation under the pure-strategy equilibrium is zero (unless subsets of size c alternate in a systematic manner from round to round). Clearly, the simulation results provide no support for the pure-strategy equilibrium. It appears that the simulation reached a state that is close to an asymmetrical mixed strategy equilibrium.

Is There "Learning-Free" Magic?

The simulation results presented in Fig. 1 suggest that the Roth–Erev model is a sufficient explanation for the "magical" coordination in the current setting. Yet, this explanation cannot account for Sundali *et al.*'s (1995, Exp. 1) observation of good coordination even in the absence of feedback (and, therefore, the possibility of learning). In Rapoport *et al.* (in press), we hypothesized that this observation may be a characteristic of Sundali *et al.*'s experimental design in which the market size changed every trial. We argued that this design may induce subjects to use cutoff strategies, and we showed that a uniform selection among cutoff strategies leads to a near equilibrium behavior in Sundali *et al.*'s task. This explanation implies that in the current "one market at a time" task, coordination is less likely to occur before the players gain experience. To test this assertion, we examined the outcome of the first period played in Condition 3 (known payoff rule). The results show no indication of coordination. In fact, more players (25 of 48) chose to enter in game $c = 4$ than in game $c = 8$ (17 of 48). These results support Rapoport *et al.*'s hypothesis and the view that coordination without learning is not a general phenomenon.

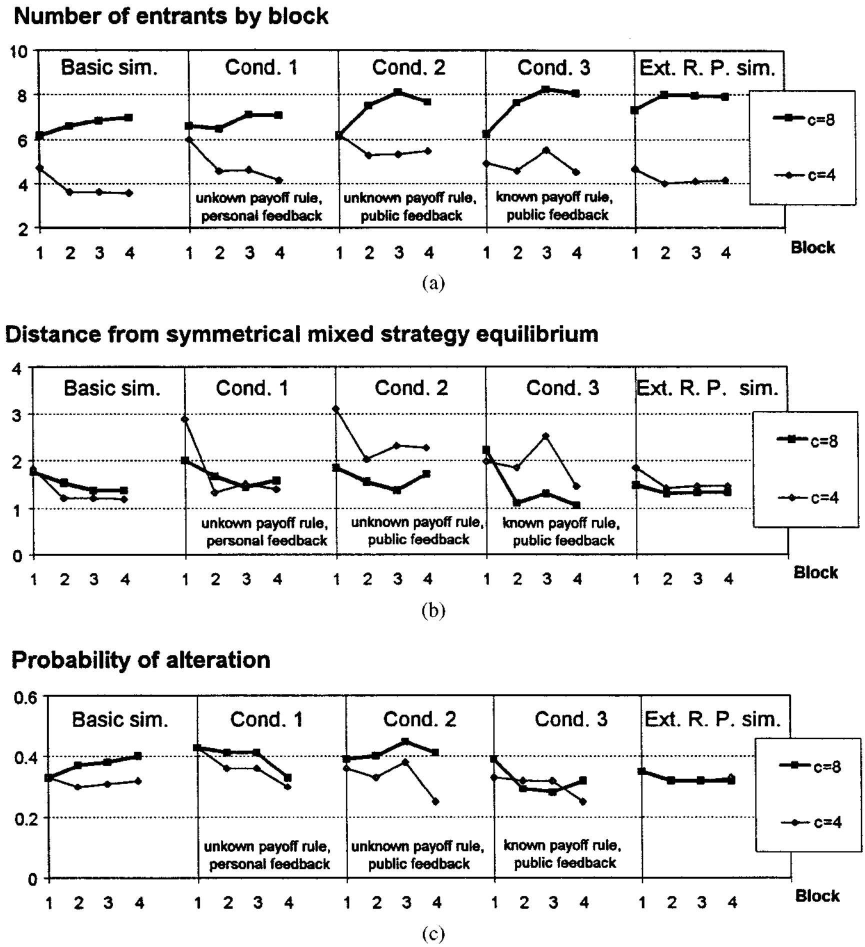

FIG. 1. Main simulation and experimental results as a function of time (four blocks of five periods each) and market capacity ($c = 4$ and $c = 8$): (a) number of entrants; (b) distance from symmetrical mixed strategy equilibrium—the absolute difference between the observed number of entrants and the expected number in symmetrical mixed strategy equilibrium; (c) probability of alternation.

The Observed Learning Processes

To assess the descriptive power of the model, we turn next to comparison of the simulated and observed results. The mean number of entrants for both games with $c = 4$ and $c = 8$ in Conditions 1, 2, and 3, are presented in columns 2, 3, and 4, respectively, of Fig. 1a. To compare the observed and simulated learning functions, the results are displayed again in terms of four blocks of five rounds each. Convergence to equilibrium is observed for all three experimental conditions; the only possible exception is game $c = 4$ in Condition 2.

The mean number of entrants were subjected to a $2 \times 2 \times 3 \times 4$ (market capacity by presentation order by information condition by block) ANOVA with repeated measures on the first and fourth factors. (We used the group as the unit of analysis and grouped the number of entrants into four blocks.) The ANOVA yielded a significant main effect due to market capacity ($F_{1,6} = 131$, $p < 0.0001$) and showed that this effect increases with time (the capacity by block interaction is significant, $F_{3,18} = 8.15$, $p < 0.005$). As shown in Fig. 1a, in all three experimental conditions the mean number of entrants in the larger market ($c = 8$) exceeded the mean number of entrants in the smaller market ($c = 4$). This result was obtained in each of the 12 groups of subjects on the final block.

In addition to the market capacity main effect and the capacity by block interaction effect, only the information condition and the order of presentation main effects were significant. The order effect ($F_{1,6} = 7.61$, $p < 0.05$) reflects transfer from the first to the second market: Over the two markets, the average number of entrants was 5.94 when the small market was presented first, and 6.35 when the larger market was presented first. As can be seen in Fig. 1a, the information effect ($F_{2,6} = 7.65$, $p < 0.05$) is due to the difference between the first condition (personal feedback) and the other two conditions (public feedback). This difference is significant ($F_{1,10} = 5.17$, $p < 0.05$), whereas the difference between Condition 3 and the first two conditions as well as all pairwise comparisons are not significant.

A comparison of the observed and simulated learning functions in Fig. 1a shows that the Roth–Erev model tracks the general trend of the aggregate results reasonably well. For both observed and simulated results, the learning function increases with experience for $c = 8$ and decreases for $c = 4$, the former function is steeper than the latter, and the major change occurs between blocks 1 and 2.

Figure 1b (columns 2, 3, and 4) displays the mean deviation from the equilibrium point. Only two main effects are significant (using the four-way repeated measure ANOVA described above): a time (block) effect ($F_{3,18} = 7.64$, $p < 0.005$) that implies learning toward the equilibrium, and a market capacity effect ($F_{1,6} = 31$, $p < 0.005$). A comparison of the simulated and observed mean results shows similar learning trends. Yet, the Roth–Erev model fails to predict the market capacity effect in Conditions 2 and 3 (this effect also violates the mixed-strategy equilibrium prediction).

Figure 1c (columns 2, 3, and 4) portrays the probability of alternation in Conditions 1, 2, and 3, respectively. The observed results are much closer to the simulation results than to either the pure- or mixed-strategy equilibrium solutions. However, in all three information conditions, the probability of alternation is seen to decrease with experience ($F_{3, 18} = 4.2$, $p < 0.05$), particularly for the game $c = 4$. In contrast, the Roth–Erev model predicts a more or less flat function for $c = 4$ and an increasing function for $c = 8$. As Fig. 1c shows, the Roth–Erev model correctly predicts a significant market capacity effect ($F_{1, 6} = 10.6$, $p < 0.05$).

Previous findings in the market entry game showed considerable individual differences in decision policies, with some subjects exhibiting hardly any alternations between the 10 iterations of the same value of c (on nonconsecutive rounds) and others exhibiting many alternations. To gain better understanding of individual differences, we computed for each subject separately the mean number of alternations over the 20 iterations of each of the 2 c values. The frequency distributions of the individual mean number of alternations for Conditions 1, 2, and 3 are displayed in panels 3, 4, and 5 of Fig. 2. The top panel of this figure shows the frequency distribution of number of alternations for players adhering to the mixed-strategy equilibrium, whereas the second panel from the top shows the frequency distribution computed over the 12×500 stat-rats. The theoretical and observed frequency distributions are shown separately for the two games with $c = 4$ and $c = 8$.

Inspection of Fig. 2 shows that, in line with previous results, the three experimental distributions are more widely spread than the ones predicted by the symmetrical mixed-strategy equilibrium solution. Although the simulated distributions are closer to the experimental distribution, they are not wide enough.

Alternative Explanations of the Information Condition Effect

A within-subject analysis was conducted to compare alternative explanations to the observed information effect. In this analysis, the predictions of the basic reinforcement model were compared to the predictions of models assuming that subjects are affected by "higher level" information. Four of the models which appear most frequently in the literature were chosen and then compared in this analysis to the Roth–Erev (RE) model: A variant of the basic Roth–Erev model assuming an *external* reference point (other players' payoffs) (RE^e), an imitation learning model (IM), a simple best reply model (BR), and a fictitious play model (FP).

At the first step of the analysis predictions were obtained for each of the decisions made by the subject, given each of the five models. The predictions of the RE model were calculated based on the cumulative reinforcements using assumption A2* (with the parameters used in the left-hand column of Fig. 1). A uniform distribution of initial propensities was assumed.

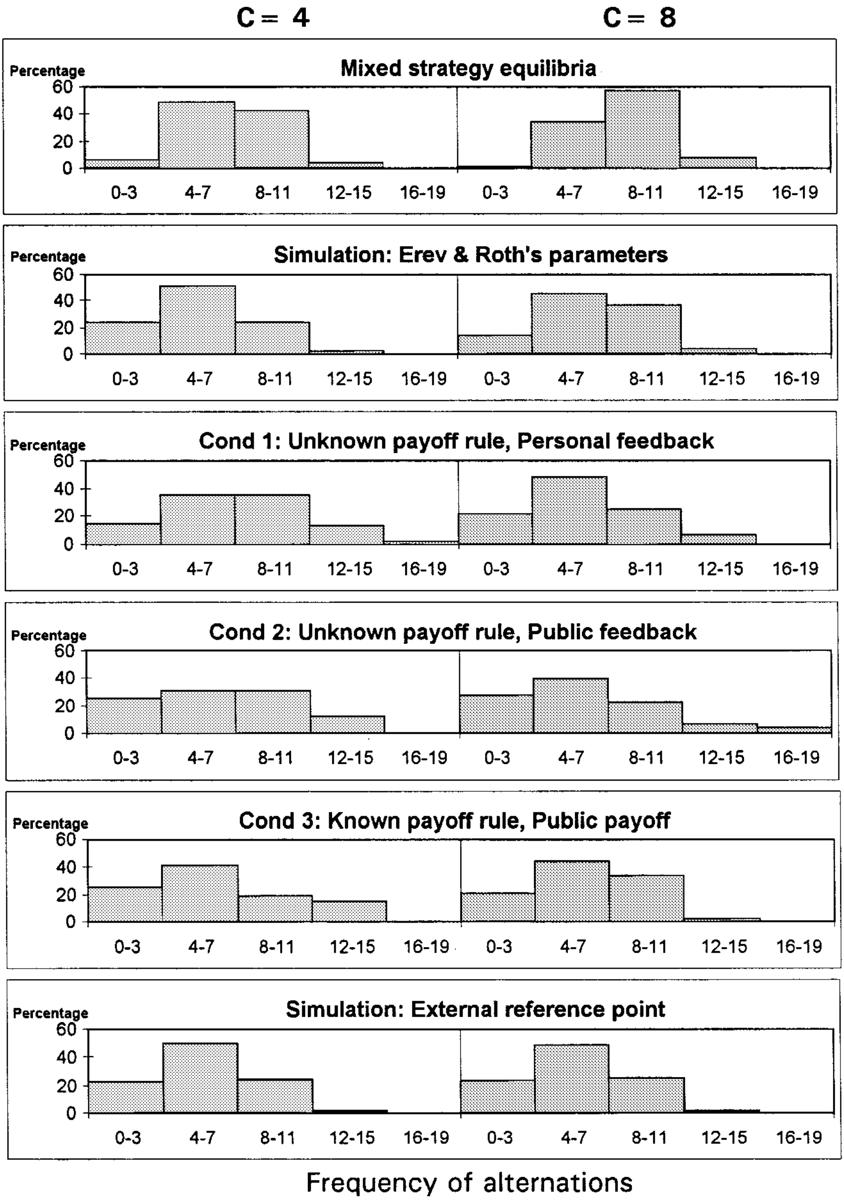

FIG. 2. Predicted and observed frequency distributions of number of alternations.

The external reference point variant of the Roth–Erev model (RE^e) considered here assumes that information concerning other player's payoff determines the player's reference point. Following a common idiom (in Hebrew), we refer to the assumption that distinguishes this model from the original RE model as the "Big Eyes" assumption. The predictions of the RE^e variant were calculated by the algorithm used for the RE model (assumption A2*) with one exception: The reference point $\rho(t)$ for player i was defined as the mean payoff of the subjects whose payoff was different from player i's payoff.

The imitation (IM) model's predictions were obtained under the assumption that in round $t + 1$ the subject imitates the decision that resulted in the highest payoff in round t. If the two strategies resulted in the same payoff, the subject is assumed to repeat his/her own strategy.

Before presenting the two remaining models, it is important to note that the experiment was designed to ensure that not all learning rules could be used by the subjects in all three experimental conditions. We shall use this aspect of the design in the evaluation of the model fitness scores, but ignore it in the calculation of these scores. That is, all statistics were computed based on the actual results. For example, we have computed the correspondence between the imitation model's predictions and actual behavior in Condition 1 even though the subject could not imitate other subjects in this condition.

The remaining two models are expectation-based learning models. Like the imitation learning model, they are deterministic in nature; they predict that the subject will enter with probability 1, if he or she expects that this decision will maximize payoff. The two models differ from each other with respect to the information that is assumed to give rise to these expectations. According to the best reply (BR) model, player i behaves in round $t + 1$ as if he or she expects that the other players will not change their behavior from round t to round $t + 1$. That is, player i is expected to choose the strategy that maximizes profit, given the other players' last choice.

Finally, according to the fictitious play (FP) model, the player's expectations are based on the assumption that all the other players decide to enter with the same fixed probabilities. Thus, the player is predicted to assess this probability and then choose the appropriate best response. Note that this rule is the natural extension of the Fictitious Play rule in two-person games (Robinson, 1951; a similar extension was studied by Cooper *et al.*, 1997).

At the second stage, fitness scores were calculated for each prediction. To facilitate comparison between the probabilistic and the deterministic models, we used the proportion of accuracy point predictions (PA) score.

TABLE I
Proportion of Accurate Individual Choices Predictions by Five Models[a]

Condition (payoff rule, feedback)	n	Proportion of accurate prediction				
		RE	RE^e	IM	BR	FP
1. (Unknown, Personal)	48	0.63	(0.63)	(0.56)	(0.58)	(0.58)
2. (Unknown, Public)	48	0.59	0.61	0.54	(0.55)	(0.57)
3. (Known, Public)	48	0.60	0.64	0.54	0.57	0.60

[a] When the information assumed to be used by the model was not available to the subjects, the relevant proportion is written in parentheses.

The PA score is equal to 1, if the subject makes the most likely choice under the model, and 0, otherwise. (When both strategies were equally likely under the model, the model's prediction was set to be the subject's choice on the previous round.)

At the final stage, the average PA scores were computed for the last 19 decisions (rounds 2–20) made by each subject in each game, and comparison statistics were calculated for each of the 48 subjects in each of the 3 information conditions.

The mean fitness scores are presented in Table I. The scores of the models that could not be used by the subjects (because they did not have the necessary information to perform the assumed computation) are shown in parentheses. Across all subjects, the RE model provides the best fit in Condition 1. In the other two conditions, the external reference point variant of the model (RE^e) outperforms the other models.

Recall that the main goal of the present analysis is to examine how the subject's behavior differs from the behavior predicted by the RE model. To address this issue, we inquire if each of the four alternative models captures a significant tendency of the subject to use information that is not expected to be used under the RE model. Thus, a difference score (D_j) between the RE fitness score and model j's fitness score (F_j) was computed for each subject. For example, if for a certain subject RE $= 0.65$ and FP $= 0.6$, then $D_{FP} = 0.05$. Under the assumption that model j describes the way subjects utilize the information concerning other subjects' behavior, D_j should be affected by the information conditions. Larger D_j values are expected when model j cannot be used (the score j is presented in parentheses).

Analysis of variance reveals that only one of the four difference score, D_{RE^e}, was affected by the model's usability ($F_{1,142} = 5.45$, $p < 0.025$). The remaining three scores were not affected by the availability of the neces-

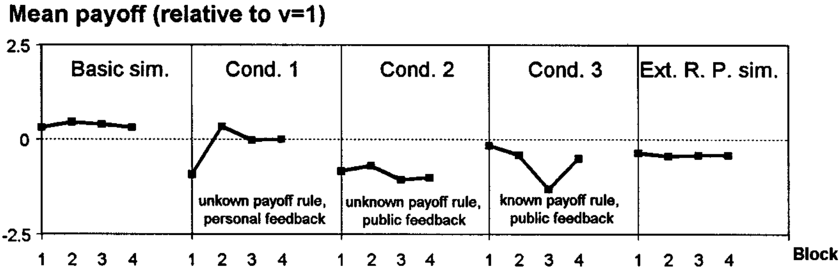

FIG. 3. Mean payoff (minus 1, the payoff for staying out) as a function of block in the format of Fig. 1.

sary information ($p > 0.05$). Similar results were obtained in an analysis using groups rather than individual subjects as the units of analysis.

Simulation of the External Reference Point Variant of the Reinforcement Learning Model

The right-hand column in Fig. 1a–c and the bottom panel in Fig. 2 display simulation results for the external reference point variant of the reinforcement learning model. In line with the within-subjects analysis, the simulation provides a better fit to the results obtained in the Conditions 2 and 3 than the RE model. Most important, the ‘‘Big Eyes’’ assumption seems to explain why public feedback increased the number of entrants.

Note that whereas both the original and the external reference point models appear to predict a convergence toward an equilibrium, the predicted efficiencies (average payoffs) are quite distinct. The larger number of entrants given the external reference point reduces efficiency. Panels 1 and 5 in Fig. 3 (average payoffs over market sizes in Fig. 1's format) illustrate this point. Panels 2, 3, and 4 in Fig. 3 show that in line with the ‘‘Big Eyes’’ assumption, in the current settings additional information impairs earnings.

Did Subjects Use Imitation- or Expectation-Based Rules?

The observation that the availability of information needed to follow an imitation or expectation learning rule did not significantly increase the proportion of decisions that can be described by the rule is slightly disturbing. This null result is inconsistent with informal information that was collected in post-experimental discussions with the subjects in which many reported that they did develop expectations and tried to maximize payoffs.

Two explanations of this inconsistency could be considered. First, in line with Nisbett and Wilson (1977), it could be argued that the subjects "report more than they can tell." That is, they think that they used expectation and imitation, whereas in fact, unconsciously simple reinforcement guided their behavior.

A second, possibly more plausible explanation is based on the hypothesis that different subjects used contradicting rules. Under this view, the weak descriptive power of the imitation and expectation models might imply that the frequency of utilization of each of these rules is roughly similar to the frequency of utilization of the opposite rules. For example, for each imitation-based decision there was a "best reply to imitation" (see Stahl and Wilson, 1995, for a similar argument). Computer simulations suggest that utilization of strategies of these types can increase the average number of entrants.

Within-group balance can be predicted under the hypothesis that imitation and expectation rules are part of the strategies considered by the players. Under the cognitive game theoretic approach of Roth and Erev, players learn to choose among these strategies in an adaptive fashion. Thus, in the present game if, for some reason, some players follow a pure strategy repeatedly, other players are predicted to learn to follow the best response to that strategy.

Under this view, negative correlations are predicted between the utilization (fitness score) of specific strategies by different group members. These negative correlations are expected to be sensitive to the availability of the information that is necessary in order to use the different strategies.

To evaluate this hypothesis, three additional scores were computed for each subject. The score O_j for player i was defined as the mean of the fitness scores of player i's 11 group members for model j. The correlations between the individual players' scores and their group averages $(r(F_j, O_j))$ are presented in Table II. In support of the present hypothesis, stronger negative correlations were observed for each of the three rules when the information needed to follow the rule was available.

For each of the three strategies, an analysis of covariance was conducted in order to test whether the linear relation between player i's score and his/her group member average score was significantly affected by the availability of the relevant information. A marginal significant effect of the availability of the information was detected for the imitation strategy ($F_{1,143} = 2.8$, $p < 0.05$, one-tail test). The effect was insignificant for the expectation strategies ($p > 0.05$). Thus, it appears that the hypothesis that imitation−expectation strategies were utilized cannot be completely rejected. Additional data are needed in order to evaluate the information effect.

TABLE II
Correlations between Individual Players' Fitness Score and Their Group Average Scores

Condition (payoff rule, feedback)	n	Statistic		
		$r(F_{IM}, O_{IM})$	$r(F_{BR}, O_{BR})$	$r(F_{FP}, O_{FP})$
1. (Unknown, Personal)	48	(−0.00)	(0.03)	(−0.08)
2. (Unknown, Public)	48	−0.22	(−0.17)	(−0.26)
3. (Known, Public)	48	−0.39	−0.26	−0.37

COMPARISON WITH PREVIOUS STUDIES

The present results suggest that the "magic" observed in previous market entry experiments can be accounted for by simple reinforcement learning processes. The "invisible hand" that quickly leads individual agents to equilibrium in these games may be similar to the "hand" that affects bar pressing decisions of rats and pigeons.

At the same time, the present results suggest that the basic Roth−Erev model may be too simple. In violation of this model, our subjects were affected by information concerning other players' payoff. The effect of the additional information can be described by an external reference point variant of the Roth−Erev model, but can also be described by the assumption that subjects use contradiction imitation−expectation based strategies.

It should be emphasized that the "Big Eyes" assumption that underlies the external reference point model is not suggested here as a general principle that describes the classification of reinforcement whenever players know other players' payoffs. It is easy to see that there are settings in which this assumption cannot capture behavior. One setting involves repeated two-person games in which players are strongly interdependent. Consider, for example, games with a Pareto efficient equilibrium and an opportunity to earn more than the other player (but less than the equilibrium outcome) by deviating from this equilibrium. Whereas the external reference point model predicts coordination failure, experimental results (see, e.g., Van Huyck *et al.*, 1990) report efficient coordination.

Another setting in which the external reference point model is inconsistent with experimental results involves asymmetrical games in which the average payoff of the "weaker" players is below the average payoffs of other players. The "Big Eyes" assumption implies that, for the weak players, all strategies will be negatively reinforced; for that reason, no learning is expected to take place. There are many known violations of this prediction. For example, consider the learning of the weak players in the ultimatum game (Roth *et al.*, 1991) and the Best Shot game (Prasnikar and Roth, 1992).

Thus, the extent to which players behave as if their reference point is affected by external outcomes appears to be context-dependent. The counterexamples presented above illustrate that at least two factors (interdependency and asymmetry) are likely to eliminate the effect of external payoffs.[4]

Another factor that is likely to affect the significance of external payoffs is the salience of this information. In Conditions 2 and 3 of the current experiment, other player's payoffs were presented to the players in the same way that their own payoffs were presented (the experimenter had stated all values). It is reasonable to hypothesize a smaller effect of external payoffs when they are not explicitly presented. To test this hypothesis and assess the robustness of the current approach, the following section examines previous market entry experiments.

Comparison with Previous Market Entry Results

The previous market entry experiments (Rapoport *et al.*, in press; Sundali *et al.*, 1995) focused on complete information conditions similar to Condition 3 of the present paper. Following each trial, subjects were informed of their personal payoff and the number of entrants. In principle, the subjects could calculate the payoff of their team members (although this information was not provided directly). Whereas these experiments were not designed to assess the effect of this information, the data can be used to evaluate the external reference point model. To achieve this goal, we compare the model proposed by Rapoport *et al.* with the external reference point variant of the Roth–Erev model.

To adapt the Roth–Erev model to their experimental task (market entry games in which the market size randomly changes from trial to trial), Rapoport *et al.* assumed that on each trial the player first selects a cutoff point and then chooses to enter if and only if the market size exceeds that point.

Recall that Rapoport *et al.* examined behavior in 20-player computerized market entry games that were played for 100 trials. The value of c was varied randomly from period to period ($c = 1, 3, \ldots, 19$). As in the present study, the payoff for entering the market was $1 + 2(c - m)$. The payoff for staying out (v) varied from group to group. It assumed the values -10 in the first group, -6 in a second, and $+6$ in the third.

The left-hand columns in Fig. 4a–c present the main experimental results of Rapoport *et al.* in 5 blocks of 20 trials (2 repetitions of each c

[4] Some support for the hypothesis that the "Big Eyes" assumption is likely to provide a good approximation of behavior given symmetry and low interdependency comes from the observation that information concerning the payoff of symmetrical players in large populations has a major effect on behavior (e.g., Bolton, 1991; Harrison and McCabe, 1996).

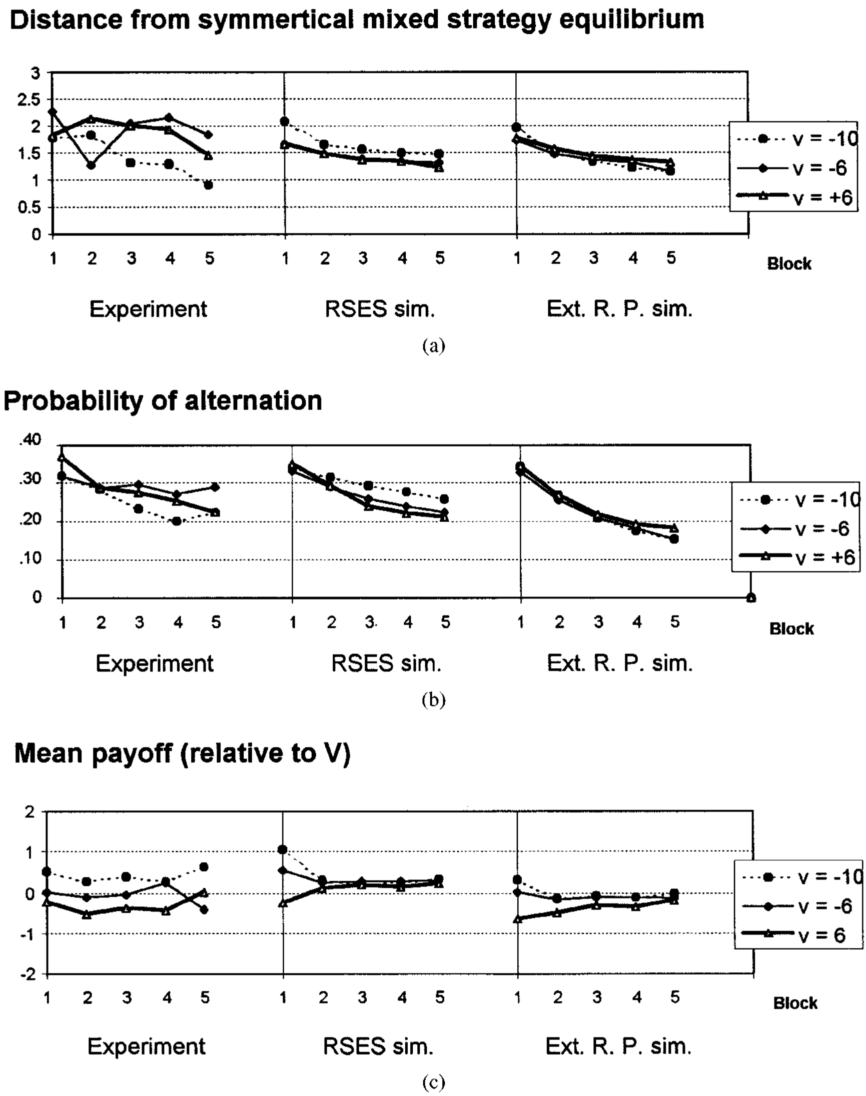

FIG. 4. Main experimental and simulation results of Rapoport *et al.* (in press) as a function of time (5 blocks of 20 periods each) and payoff condition: (a) distance from symmetrical mixed strategy equilibrium—the absolute difference between the observed number of entrants and the expected number in symmetrical mixed strategy equilibrium; (b) probability of alternation; (c) mean payoff (minus V, the payoff for staying out).

value in each block). The main experimental trends are consistent with the current findings: a very quick convergence to the mixed-strategy equilibrium number of entrants, a decrease in the number of alternations over time, and relatively flat payoff curves.

The central panels in Fig. 4–c display the predictions of the reinforcement learning model proposed by Rapoport *et al.* (in press), referred to as the RSES model. This model assumes that subjects set their initial reference point at $\rho(1) = v$ and then update it based on their personal payoff (in line with the basic Roth–Erev model). The right-hand panel of Fig. 4 displays the prediction of the external reference point variant of the model. The two variants of the model provide similar predictions of the distance from equilibrium and the probability of alternation (Fig. 4a and 4b). Both models capture the major trends with regard to these statistics. A larger difference between the models is observed in Fig. 4c. The external reference point model predicts lower expected payoffs. The experimental results seem to fall between the two predicted curves.

Similar findings are displayed in Fig. 5a–c, which present the main results in Experiment 2 of Sundali *et al.* (1995). This experiment had a design similar to Rapoport *et al.*'s study with the exception that $v = 1$ in all three groups. Sundali *et al.*'s results are consistent with Rapoport *et al.*'s results. The distance from equilibrium and the number of alternation functions are captured by the two variants of the reinforcement learning model, whereas the payoff functions fall between the predictions of the two variants.

A quantitative assessment of goodness of fit can be obtained by comparing the distance between the different functions. Mean squared deviation scores (MSD) were calculated for each model. The models' MSD scores were computed as the average squared distance between the model's curve and the three group curves in each experiment. Thus, each score is an average of 15 squared differences (3 groups $\times$ 5 points in time) in each experiment. Low MSD scores reflect better fit.

Table III presents the MSD scores. It shows that over the two experiments the "Big Eyes" assumption improves the fit of the entrants (distance from equilibrium) and the profit curves, but impairs the fit of the alternation curve.

To evaluate the significance of the MSD scores, it is useful to compare the models' scores with a between-group scores. The bottom row in Table III presents the MSD scores between the curves of the three groups in Sundali *et al.*'s experiment. Note that all the three "between-groups" scores are larger than the models' scores for the relevant experiment. These results imply that, on the average, each experimental group was closer to each of the two models than to the other two groups.

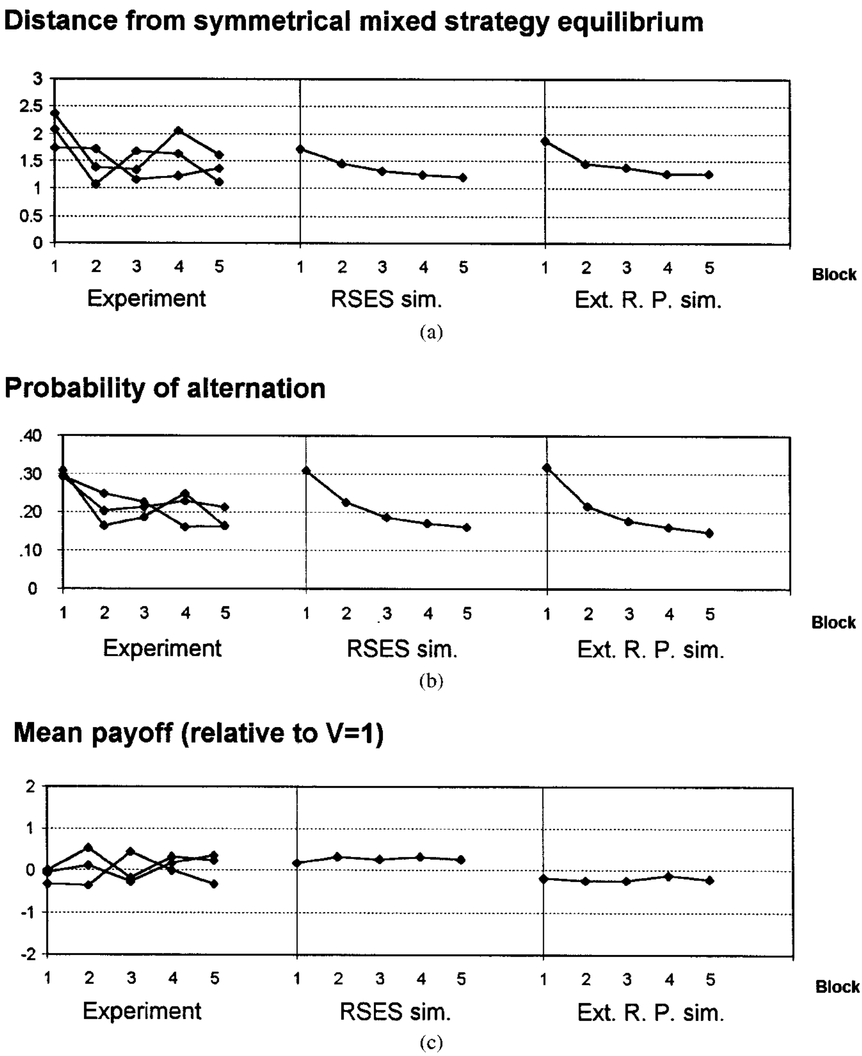

FIG. 5. Main experimental and simulation results of Sundali *et al.* (1995) as a function of time (5 blocks of 20 periods each): (a) distance from symmetrical mixed strategy equilibrium —the absolute difference between the observed number of entrants and the expected number in symmetrical mixed strategy equilibrium; (b) probability of alternation; (c) mean payoff (minus 1, the payoff for staying out).

TABLE III
Mean Squared Difference (MSD) Scores for the Functions in Figs. 4 and 5 (Rapoport *et al.*, in Press; Sundali *et al.*, 1995).

		MSD for		
Model/curves:	Experiment:	Distance from equilibrium	Probability of alternation	Average payoff
Rapoport *et al.* variant (RSES)	Rapoport *et al.*	0.239	0.0014	0.175
	Sundali *et al.*	0.123	0.0013	0.129
	Average	0.181	0.0014	0.152
The Ext. R. P. variant	Rapoport *et al.*	0.215	0.0036	0.105
	Sundali *et al.*	0.109	0.0018	0.141
	Average	0.162	0.0027	0.123
Between groups	Sundali *et al.*	0.203	0.0020	0.210

In summary, the present analyses suggest that both variants of the Roth−Erev model provide reasonable approximation to the aggregate results of Rapoport *et al*. and Sundali *et al*. The observed mean results fall somewhere between the predictions of the two variants of the model. The comparison of these results with Conditions 2 and 3 of the current experiment supports the hypothesis that the salience of external payoffs affects their effect. Rapoport *et al*.'s and Sundali *et al*.'s results are best described by a model that assumes that the players' reference point is sensitive to both internal and external payoffs. Clearly, additional research is needed in order to improve our understanding of the conditions that affect the determination of the reference point.

Comparison with Similar n-Person Games

The fast learning observed in the current setting is consistent with the results of repeated coordination games (Ochs, 1990; Meyer *et al*., 1992)[5], but differs from the slow learning typically observed in *n*-person experimental games with Pareto deficient equilibria. One example is the team game studied by Bornstein *et al*. (1994) following Palfrey and Rosenthal

[5] Meyer *et al*. observed quick convergence to a near mixed strategy equilibrium (on aggregate relative frequency, but not at the individual level) and a suggestion for a pure equilibrium play by experienced subjects. Ochs' subjects (in the zero turnover condition) were closer to a pure strategy equilibrium. The relative advantage of the pure strategy equilibrium in Ochs' study is consistent with the Roth−Erev model: in Ochs' multiple small markets game exploration is costly and only rarely reinforced.

(1983). Whereas this game has a unique pure-strategy equilibrium (in which 100% investment rate is predicted), Bornstein *et al.* found 50% investment rate and practically no learning trend in 20 rounds. Bornstein *et al.* found that their results are consistent with the Roth−Erev model. In their game, this simple reinforcement learning rule predicts that behavior will be ''locked'' between two mixed-strategy equilibria (see Palfrey and Rosenthal, 1983), and as a result the payoff function is flat and no learning is observed.

Slow learning process is also predicted by the Roth−Erev model in step-level public goods problems that have contradicting multiple equilibria. Figure 6 demonstrates that this prediction is consistent with experimental results. The left-hand panel shows the proportion of contribution in five blocks of five trials each in the three conditions studied by Rapoport and Eshed-Levy (1989). The ''Fear & Greed'' condition is a five-person step-level public goods game in which the contribution cost is 2 units, and the public good is a reward of 5 units that each player gets if at least three players contribute. The ''No Fear'' condition is a variant of the first condition in which contributors do not lose their contribution if the public good is not provided. Finally, in the ''No Greed'' condition all players are charged for the contribution if the public good is obtained. After each choice, the subjects received feedback about their personal payoff. Information about other players' decisions and payoff was not provided. Thus, the external reference point assumption does not apply. As can be seen in the right-hand side of Fig. 6, the Roth−Erev model (with the parameters used above) captures the slow experimental learning trends.

These results support the assertion that the difference between the current setting and situations in which subjects appear to be a slow

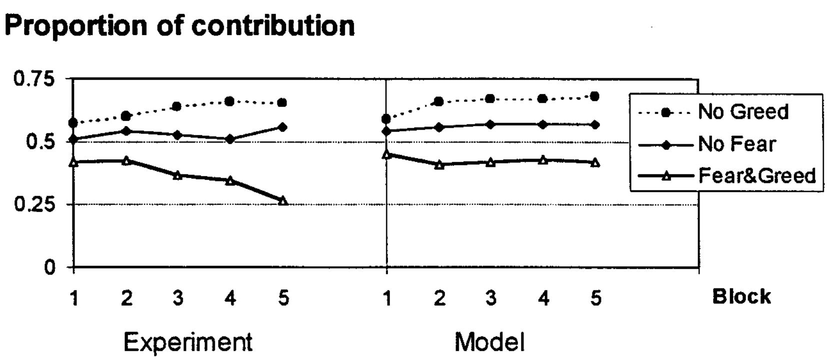

FIG. 6. Proportion of contribution to the public good in the experiment and simulation of Rapoport and Eshed-Levy (1989) as a function of time (five blocks of five periods each).

learners may not be a result of substantially different learning processes. Rather, it is possible that the same learning process (that can be approximated by the Roth−Erev model) leads to fast convergence to equilibrium in some games and to slow convergence in others.

CONCLUSION

In line with recent research (e.g., Roth and Erev, 1995; Erev and Roth, in press; Bornstein *et al.*, 1994; Rapoport *et al.*, 1997; Rapoport *et al.*, in press; Ochs, 1995a; Erev *et al.*, 1995; Slonim and Roth, in press; Gilat *et al.*, 1997; Camerer and Ho, in press; Tang, 1996; Mookherjee and Sopher, 1997; Nagel and Tang, in press), the present paper demonstrates that simple reinforcement learning models provide a good approximation of aggregate behavior in a wide setting of experimental games.

The present paper improves our understanding of the value of reinforcement learning processes in four important ways. First, it illustrates that reinforcement learning (and observed learning) can lead toward an asymmetrical mixed-strategy equilibrium. Second, it shows that the Roth−Erev quantification provides better approximation of behavior when the players do not know other players' payoffs. Third, it demonstrates that, in market entry games, the effect of information concerning other players' payoff can be modeled by the assumption that this information affects the players' reference point. Finally, the present paper reveals that even minor quantitative differences in the assumed learning process can have a nontrivial effect on the market efficiency. Thus, the fact that good approximation of behavior can be achieved by simple reinforcement learning models does not imply that the details of the learning process can be ignored.

REFERENCES

Bolton, G. (1991). "A Comparative Model of Bargaining: Theory and Evidence," *Amer. Econ. Rev.* **81**, 1096−1136.

Bornstein, G., Erev, I., and Goren, H. (1994). "Learning Processes and Reciprocity in Intergroup Conflicts," *J. Conflict Resolution* **38**, 690−707.

Bush, R., and Mosteller, F. (1955). *Stochastic Models for Learning*. New York: Wiley.

Camerer, C. (1990). "Behavioral Game Theory," in *Insights in Decision Making: A Tribute to Hillel J. Einhorn* (R. Hogarth, Ed.). Chicago: Univ. of Chicago Press.

Camerer, C., and Ho, T. (in press). "Experience-Weighted Attraction Learning in Games: Estimates from Weak Link Games," in *Games and Human Behavior: Essays in Honor of Amnon Rapoport's 60th Birthday* (D. V. Budescu, I. Erev, and R. Zwick, Eds.).

Cooper, D., Garvin, S., and Kagel, J. (1997). "Signaling and Adaptive Learning in an Entry Limit Pricing Game," *Rand J. Econ.* **28**, 662−683.

Crawford, V. P. (1995). "Theory and Experiment in the Analysis of Strategic Behavior," unpublished paper presented at the Seventh World Congress of the Econometric Society, Tokyo, August 1995.

Erev, I., Gopher, D., Itkin, R., and Greenshpan, Y. (1995). "Toward a Generalization of Signal Detection Theory to *N*-Person Games: The Example of Two-Person Safety Problem," *J. Math. Psychology* **39**, 360–375.

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

Erev, I., and Roth, A. (in press). "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *Amer. Econ. Rev.*

Gary-Bobo, R. (1990). "On the Existence of Equilibrium Points in a Class of Asymmetric Market Entry Games," *Games Econ. Behav.* **2**, 239–246.

Gilat, S., Meyer, J., Erev, I., and Gopher, D. (1997). "Beyond Bayes' Theorem: The Effect of Base Rate Information in Consensus Games," *J. Experimental Psychology: Appl.* **3**, 83–104.

Harley, C. B. (1981). "Learning the Evolutionary Stable Strategy," *J. Theoret. Biol.* **89**, 611–633.

Harrison, G. W., and McCabe, K. A. (1996). "Expectations and Fairness in a Simple Bargaining Experiment," *Int. J. Game Theory* **25**, 303–327.

Herrnstein, R. J. (1961). "Relative and Absolute Strength of Response as a Function of Frequency of Reinforcement," *J. Experimental Anal. Behav.*, **4**, 267–272.

Herrnstein, R. J. (1970). "On the Law of Effect," *J. Experimental Anal. Behav.*, **13**, 243–266.

Kagel, J. H., and Roth, A. E. (1995). *Handbook of Experimental Economics*. Princeton, New Jersey: Princeton Univ. Press.

Kahneman, D. (1988). "Experimental Economics: A Psychological Perspective," in *Bounded Rational Behavior in Experimental Games and Markets* (R. Tietz, W. Albers, and R. Selten, Eds.). Berlin: Springer-Verlag.

Meyer, D., Van Huyck, J., Battalio, R., and Saving, T. (1992). "History's Role in Coordinating Decentralized Allocation Decisions: Laboratory Evidence on Repeated Binary Allocation Games," *J. Polit. Econ.* **100**, 292–316.

Mookherjee, D., and Sopher, B. (1997). "Learning and Decision Costs in Experimental Constant Sum Games," *Games Econ. Behav.* **19**, 97–132.

Nagel, R., and Tang, F. (in press). "Experimental Results on the Centipede Game in Normal Form: An Investigation of Learning," *J. Math. Psychology*.

Nisbett, R. E., and Wilson, T. D. (1977). "Telling More than We Can Know: Verbal Reports on Mental Processes," *Psychological Rev.* **84**, 231–259.

Ochs, J. (1990). "The Coordination Problem in Decentralized Markets: An Experiment," *Quart. J. Econ.* **105**, 545–559.

Ochs, J. (1995a). "Simple Games with Unique Mixed Strategy Equilibrium: An Experimental Study," *Games Econ. Behav.* **10**, 202–217.

Ochs, J. (1995b). "Coordination Problems," in *Handbook of Experimental Economics* (J. Kagel and A. E. Roth, Eds.). Princeton: Princeton Univ. Press.

Palfrey, T., and Rosenthal, H. (1983). "A Strategic Calculus of Voting," *Public Choice* **41**, 7–53.

Prasnikar, V., and Roth, A. E. (1992). "Considerations of Fairness and Strategy: Experimental Data From Sequential Games," *Quart. J. Econ.* **107**, 865–888.

Rapoport, A. (1995). "Individual Strategies in a Market-Entry Game," *Group Decision and Negotiation* **4**, 117–133.

Rapoport, A., and Eshed-Levy, D. (1989). "Provision of Step-Level Public Goods: Effect of Greed and Fear of being Gypped," *Organizational Behavior and Human Decision Processes* **44**, 325–344.

Rapoport, A., Erev, I., Abraham, E. V., and Olson, D. E. (1997). "Randomization and Adaptive Learning in a Simplified Poker Game," *Organizational Behavior and Human Decision Processes* **69**, 31–49.

Rapoport, A., Seale, D. A., Erev, I., and Sundali, J. A. (in press). "Coordination Success in Market Entry Games: Tests of Equilibrium and Adaptive Learning Models," *Management Sci*.

Rapoport, An., Guyer, M., and Gordon, D. (1976). *The* 2×2 *Game*. Ann Arbor: Univ. of Michigan Press.

Robinson, J. (1951). "An Iterative Method of Solving a Game," *Ann. Math*. **54**, 296–301.

Roth, A. E. (1995). "Introduction to Experimental Economics," in *Handbook of Experimental Economics* (J. Kagel and A. E. Roth, Eds.). Princeton, New Jersey: Princeton Univ. Press.

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

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., and Zamir, S. (1991). "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *Amer. Econ. Rev*. **81**, 1068–1095.

Selten, R., and Güth, W. (1982). "Equilibrium Point Selection in a Class of Market Entry Games," in *Games, Economic Dynamics, and Time Series Analysis* (M. Diestler, E. Fürst, and G. Schwadiauer, Eds.). Wien-Würzburg: Physica-Verlag.

Slonim, R., and Roth, A. E. (in press). "Financial Incentives and Learning in Ultimatum and Market Games: An Experiment in the Slovak Republic," *Econometrica*.

Stahl, D., and Wilson, P. (1995). "On Players' Models of Other Players: Theory and Experimental Evidence," *Games Econ. Behav*. **10**, 218–254.

Sundali, J. A., Rapoport, A., and Seale, D. A. (1995). "Coordination in Market Entry Games with Symmetric Players," *Organizational Behav. Human Decision Processes* **64**, 203–218.

Tang, F. (1996). "Anticipatory Learning in Two-Person Games: An Experimental Study," Discussion Paper B-363. University of Bonn.

Thorndike, E. L. (1898). "Animal Intelligence: An Experimental Study of the Associative Processes in Animals," *Psychological Monographs* **2**.

Van Huyck, J., Battalio, R., and Beil, R. (1990). "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure," *Amer. Econ. Rev*. **80**, 234–248.