GAMES AND ECONOMIC BEHAVIOR **19,** 97–132 (1997)
ARTICLE NO. GA970540

Learning and Decision Costs in Experimental Constant Sum Games*

Dilip Mookherjee

Department of Economics, Boston University, 270 Bay State Road, Boston, Massachusetts 02215, and Indian Statistical Institute, 7 S.J.S. Sansanwal Marg, New Delhi 110016, India

and

Barry Sopher

Department of Economics, Rutgers University, New Brunswick, New Jersey 08903

Received May 11, 1994

Subjects played strategically similar 4×4 and 6×6 constant sum games under varying payoff scales. Substantial divergences from equilibrium predictions were exhibited. The dynamic pattern of play is best explained by a stimulus learning model whereby players allocate weight to different actions according to their relative (time average) payoff experience in past plays. The results do not provide much support for the hypothesis that players select best responses to beliefs about opponent play based on observed choice frequencies in past plays, modified by random errors or preference shocks. *Journal of Economic Literature* Classification Numbers: C72, C92.

1. INTRODUCTION

Recent experimental work has shown that subjects are able to play essentially as predicted by Nash equilibrium in some simple games, such as a matching pennies game, but that in larger constant-sum games, substan-

* For useful comments and suggestions the authors thank Saumya Banerjee, Tom Palfrey, Bob Rosenthal, two anonymous referees, and participants at the Experimental Games Workshop at the 1993 Stony Brook Summer Institute in Game Theory. We also thank Eli Berman and Choon-Geol Moon for their advice concerning model selection methods in multinomial logits. The experiment was made possible by the financial support of the Rutgers Research Council, the organizational support of Kaushik Basu and Badal Mukhopadhyay at the Delhi School of Economics, and the assistance of Shanto Ghosh and Gora Gangopadhyay in conducting the experiment, to all of whom we are very grateful. Barry Sopher also thanks the Indian Statistical Institute, New Delhi, for their hospitality.

0899-8256/97 $25.00

tial deviations from equilibrium predictions can be expected (Brown and Rosenthal, 1990; McCabe et al., 1993; Mookherjee and Sopher, 1994; and Rapoport and Boebel, 1992). Some of these studies have found that experimental outcomes are not significantly altered by increasing the scale of payoffs. These findings suggest that increased cognitive complexity, rather than insufficient monetary motivation, is responsible for the failure of minmax or Nash equilibrium predictions. Moreover, they raise the question of what kinds of behavioral models can account for the nature of play observed in games where the hypothesis of Nash equilibrium behavior appears inadequate. Do players learn from experience? If so, what kind of learning models can account for observed patterns of play? How do the learning heuristics vary with the environment, such as the level of complexity or scale of payoffs?

This paper reports the results of an experiment designed to analyze these questions systematically. We study two versions of constant sum games (repeated 40 times) similar to those studied previously by Rapoport and Boebel (1992): one is a 4×4 and the other is a 6×6 game, which are strategically similar. The 6×6 game is essentially the 5×5 Rapoport–Boebel game, with the addition of a dominated action. Both games have essentially the same (unique) mixed strategy equilibrium. The 4×4 game collapses three of the (undominated) actions in the 6×6 game into a single one. The two games thus differ essentially with respect to the number of choices available to each player, representing one dimension of the level of cognitive complexity of the game. Moreover, each game is replicated at two different scales of payoff. Hence both the dimension of the game and the payoff levels are used as treatment variables. The focus of the analysis is to differentiate amongst different adaptive learning rules used by players and to identify systematic differences, if any, in the use of such rules as dimension and payoff are varied.

Learning rules are represented in the form of multinomial logit equations predicting the likelihood of players selecting different actions, depending on different variables representing outcomes of past experience. This approach generalizes the notion of a *quantal response equilibrium* (*QRE*) proposed recently by McKelvey and Palfrey (1993) in order to explain divergences from Nash equilibrium predictions in diverse experiments. Our approach augments the QRE notion by introducing the possibility of players learning from past experience, instead of simultaneously playing best (quantal) responses to one another: It may thus be referred to as *quantal response learning* (*QRL*). The QRL formulation is flexible enough to encompass a wide range of learning models discussed in the literature. For instance, besides models that involve no learning (such as minmax or QRE), models of fictitious play learning (see Shapley, 1964; Fudenberg and Kreps, 1993), Cournot *tatonnement*, Bayesian learning (see

Jordan, 1991, 1993), as well as psychological ‘‘stimulus’’ learning (see Bush and Mosteller, 1955; Suppes and Atkinson, 1960; Selten and Stoecker, 1986; Bendor *et al*., 1993; Roth and Erev, 1995), all form special cases. The model can either be interpreted within a utility-maximizing framework where preferences are subject to random shocks, or within a behavioral framework directly predicting choice probabilities as a function of different experiential variables. The formulation allows us to test for any one of the various alternative hypotheses, jointly with the multinomial logit formulation.

Briefly, our main results are as follows. There is substantial deviation from minmax play in all versions, either in terms of overall choice frequencies or in terms of the behavior of subjects. These deviations do not seem to change materially with a change in the scale of payoffs, though they do differ significantly across games of differing dimension. There is clear evidence that players learn from experience. With a single exception, for all categories of players and for all the games, the dynamics of play is best represented by a simple ‘‘stimulus learning’’ or ‘‘naive Bayesian’’ model where players allocate probability weight across different actions according to the relative values of (simple time) averages of payoffs achieved from different actions in past plays. Models of players selecting best responses to beliefs about opponent play which are based on empirical choice frequencies in past plays (such as fictitious play or the Cournot hypothesis) fit the data less well and generate coefficients violating corresponding *a priori* restrictions. Moreover, the evidence also suggests that players do not weigh the experience of more recent plays more heavily than earlier plays: if anything, the outcome of early play appears to exercise a strong influence throughout the course of the game.

Nevertheless, the estimated parameters of the learning model appear to fluctuate considerably across different games, as well as between early and later stages of any given game. There appear to be no clear patterns with respect to varying the dimension or payoff scale: hence no support is obtained for theories of ‘‘decision cost’’ where the level of sophistication of players is presumed to vary with the level of complexity and the monetary stakes (Cox, *et al*., 1990; Smith and Walker, 1993).

The paper is organized as follows. Section 2 describes the nature of the games studied and the experimental design. Section 3 introduces quantal response learning models and explains various interpretations and special cases of these. It also explains the nature of *a priori* restrictions that would be associated with specific models, which can be used to assess their empirical plausibility. Section 4 contains the empirical results. Finally, Section 5 summarizes the main findings and relates them to existing literature.

2. EXPERIMENTAL DESIGN AND PROCEDURES

The games studied in the experiment are shown below. There are two different payoff matrices showing the payoffs of the row player. A "W" denotes a win, and an "L" denotes a loss. The column player receives an L when the row player receives a W, and vice versa. In the 4×4 game, "#" refers to a lottery where W results with probability $\frac{1}{3}$, and "&" is one where W results with probability $\frac{2}{3}$. In Games 1 and 2, W represents a payment of 5 rupees to the winner and 0 to the loser. In games 3 and 4, the scale of payoffs is doubled: W refers to a payment of 10 rupees to the winner, and 0 to the loser.

Row Player Payoff Matrix for Games 1 and 3

Column Player Choice:	1	2	3	4
Row Player Choice:				
1	W	L	L	W
2	L	L	W	W
3	L	W	#	#
4	L	L	&	W

Row Player Payoff Matrix for Games 2 and 4

Column Player Choice: 1	1	2	3	4	5	6
Row Player Choice:						
1	W	L	L	L	L	W
2	L	L	W	W	W	W
3	L	W	L	L	W	L
4	L	W	W	L	L	L
5	L	W	L	W	L	W
6	L	L	W	L	W	W

Note that the 4×4 game is simply the 6×6 game with actions 3, 4, and 5 for each player collapsed into a single action (action 3). In both games, the last action is dominated (by the second for the row player, and the first for the column player). If this action is eliminated, the 6×6 game reduces to the game studied by Rapoport and Boebel (1992). The decision problem for any player is similar across the 4×4 and 6×6 games, if this player expects that the other will select the third, fourth, and fifth actions with equal probability, as is predicted by the minmax hypothesis. The 6×6 game has a unique mixed strategy equilibrium which is symmetric (despite the fact that the game is not symmetric): it entails choosing the first action with probability $\frac{3}{8}$, the second with probability $\frac{1}{4}$, and the remaining three with probability $\frac{1}{8}$ each. Under the minmax hypothesis, this game is strategically equivalent to the 4×4 game, whose unique equilibrium involves choice of the first and third actions with probability $\frac{3}{8}$ each, and the second with probability $\frac{1}{4}$.

We choose to study constant-sum games for a variety of reasons. First, Nash equilibrium corresponds to minmax play: in this context game theory makes some "confident" predictions. In the games we study, there is a unique mixed strategy equilibrium. Second, the equilibrium is typically in mixed strategies. This ensures that there is no scope for players to settle down early in the game to a pure strategy, providing experimenters with sufficient variability in the pattern of play to estimate and discriminate between different behavioral hypotheses. Third, the predictions of the minmax theory are independent of levels of risk aversion over monetary gambles of the players. Fourth, the constant-sum character of the games eliminates possible cooperative motives for not playing according to equilibrium theory predictions.

Constant-sum games of smaller dimension were not selected for a variety of reasons. First, previous experiments (such as McCabe *et al.*, 1993, and Mookherjee and Sopher, 1994) found that when players are fully informed about opponents' choices and payoffs the minmax hypothesis receives substantial support for 2×2 games, whereas 4×4 or 5×5 games do not (Brown and Rosenthal, 1990; Rapoport and Boebel, 1992). Given our interest in possible divergences from minmax play, this dictated selection of a dimension where there is a reasonable chance that such patterns would be observed. Morever, the games were chosen to be similar to that in the Rapoport–Boebel experiment to ensure comparability of results. Second, the contrast between the results of earlier experiments invites a closer examination of the effects of varying the dimension (i.e., number of actions available to each player) of constant-sum games. This aspect of the structure of a game can be thought to represent one aspect of the level of cognitive complexity of the game, which may have a significant effect on the way subjects actually play the game. For one thing, players have to digest a larger amount of information as represented in the payoff matrix and select from a larger number of alternative actions, in a game of higher dimension. To the extent that players try to guess the behavior of their opponents, they have to predict a larger number of choice probabilities. The objective of examining the sole effect of varying the dimension of the game explains why the 4×4 and 6×6 games were selected to be comparable, in the sense that they have essentially the same minmax solution, with some of the strategies in the larger game collapsed into a single strategy in the smaller one.

Some explanation for inclusion of the dominated action is also necessary. Our main motive is that plays of such actions provide evidence enabling one to discriminate between models of ‘‘stimulus’’ and ‘‘belief’’ learning (Selten, 1978, 1991). In the former, players are viewed as not pursuing optimizing behavior of any sort: instead they are presumed to discriminate between different actions on the basis of relative payoffs experienced from them in the past. In the latter, players are presumed to select best responses to their beliefs regarding opponent’s play, accumulated on the basis of information concerning past choices of the latter. The former hypothesis is consistent with significant plays of the dominated strategy, especially at early stages of the game, with later frequencies depending on the payoffs obtained thereby, relative to those associated with other choices.

In contrast, belief learning hypotheses can explain such plays only via ‘‘mistakes’’ or introduction of random shocks to payoff functions. In some contexts, such as Prisoners Dilemma games or ‘‘centipede’’ games where dominated actions correspond to cooperative outcomes, such shocks may be interpreted as reflecting the possible presence of altruistic motives (see Kreps *et al.*, 1982; McKelvey and Palfrey, 1992). Players have a strong incentive to select the dominated strategy at early stages of such games in an attempt to persuade their opponents to believe that they are of the altruistic type. In our context of a constant-sum game, no such incentive exists. Declining frequencies of dominated action choices can be reconciled with a belief learning model only via an extraneous assumption that the random shocks decline in intensity as play progresses. It could conceivably be argued that such an assumption represents some form of learning. The important point, however, is that such a learning process is not explicitly modeled and is at any rate quite orthogonal to the ‘‘belief learning’’ process. Moreover, to the extent that such shocks are interpreted as the result of ‘‘mistakes’’ or ‘‘lack of attention,’’ one would expect that their intensity would depend on how costly such lapses were. As in a standard ‘‘decision cost’’ model (Cox *et al.*, 1990; Smith and Walker, 1993), one would expect them to be less frequent when the monetary stakes are higher—a prediction which can be tested against the data.

2.1. *Procedures*

The experiment was conducted in January 1993 at the Delhi School of Economics. The subjects were master's level students in economics. The four games described above were played with different subjects, all of whom were not familiar with game theory and had no previous experience participating in game experiments. Morever, the fact that similar experiments have rarely been conducted at Indian schools or universities implied that the students would not have even known others with experience in similar experiments. Each of us separately conducted two games each in different classrooms, of differing dimension and payoff scale. Participants were randomly allocated into the different treatment groups. For each game, 20 subjects (forming 10 pairs) played against a single anonymous opponent for 40 periods, as either a row and column player. Each player repeatedly played against the same opponent, a context in which many models of learning (such as fictitious play) have been studied. Since there is no scope for cooperation, the possibility of implicit collusion in the game was not present.

Some distinctive features of our experimental setting were the following. First, levels of motivation were high, judging by levels of participation and the monetary rewards. There was a marked excess supply of subjects. Total earnings ranged from 100 to 300 rupees, averaging about 150 rupees, for an experiment lasting between two and three hours. These represented substantial monetary stakes for the typical subject (e.g., full monthly room and board in a student dorm at the time was about 600 rupees). Second, the procedure followed allowed considerable time and scope to subjects between stages of the game to reflect on their choice problem and the history of past plays.[1] Finally, the game was terminated at the end of 40 rounds to prevent player fatigue or boredom from setting in. Thus subjects played the game less frequently compared to the O'Neill (1987) or Rapoport–Boebel experiments, while on the other hand a larger number of subjects participated in our experiment.[2]

[1] Each subject was asked to mark his or her choice on a sheet of paper. Subjects knew they were playing against some other person in the same room, but could not identify their opponent; pairs were selected by a lottery at the outset of the experiment. At the end of each round, entry sheets were collected manually, the choice of the opponent was marked on each sheet, and then returned manually. No time pressure was imposed on the amount of time subjects had available to mark their choices. Successive stages were thus marked by gaps of between three to five minutes. Moreover, each player had a sheet recording the history of past plays in front of them while making their choice.

[2] In O'Neill's experiment, 25 pairs of subjects played 105 times each, while in Rapoport and Boebel (1992) 10 pairs played 240 times each; both experiments were on a face-to-face basis. Moreover, in the Rapoport–Boebel experiment, each player switched roles from row to column player (as well as the opponent) halfway through the experiment.

As with any experiment, voluntary participation of subjects implied that selection for the experiment was not random. The fact that subjects were chosen from the Masters program in economics at a single university also implies that the subjects were not selected from a very large population. However, these do not seem to us to be significant problems since our objective is to examine the behavior of the participants in the experiment, rather than to estimate relevant parameters for a larger population. We also believe that anonymity problems were insignificant, given the facts that (a) the subjects had no prior training in game theory, or shared experience in game experiments, that might have led them to form prior expectations concerning the strategy to be used by a randomly drawn member of the participant group, and (b) each player was involved in a constant-sum game with one other player, in a room containing twenty players, and was not aware of the specific identity of his or her opponent.

An appendix contains illustrative instructions for a 4×4 and a 6×6 game, as well as the detailed procedures followed in conducting the experiment. Data from the experiment are available upon request from the authors.

3. QUANTAL RESPONSE LEARNING MODELS

In this section we describe the class of learning models estimated from the data and the nature of the inferences and evaluative criteria used. The basic approach may be termed *quantal response learning*, which is a "learning" version of the QRE notion of McKelvey and Palfrey (1993). The approach also draws on the multinomial logit framework commonly used in econometric models of discrete choice (Amemiya, 1981; McFadden, 1976, 1981; Maddala, 1986). We suppose that each player selects from different action options available according to their perceived payoff characteristics. Specifically, let Π_{lt} denote a measure of "perceived payoff" from action a_l at stage t of the game. Then the probability that action a_i is selected at stage t is given by

$$p_{it} = \frac{\exp_{it}}{\sum_l \exp_{lt}}, \tag{1}$$

where

$$\exp_{lt} = \exp\left(\sum_j \alpha_{lj} \Pi_{jt} \right) \tag{2}$$

may be viewed as a "score" assigned to action a_l at t, based on a linear combination of payoffs from the different actions as perceived at t. Since a constant displacement of all coefficients will leave choice probabilities unchanged, one commonly normalizes by setting $\alpha_{nt} = 0$ all t, for some base action a_n. With this normalization, (1) reduces to

$$p_{nt} = \frac{1}{1 + \sum_{l \neq n} \exp_{lt}}$$

and for all other actions to

$$p_{it} = \frac{\exp_{it}}{1 + \sum_{l \neq n} \exp_{lt}},$$

implying that

$$\log\left(\frac{p_{it}}{p_{nt}}\right) = \sum_j \alpha_{lj} \Pi_{jt}. \tag{3}$$

Hence the coefficient α_{lj} can be interpreted as the proportionate change in the odds of choice l *relative* to the base choice a_n, given a unit increase in the perceived payoff from action a_j. This also implies that

$$\frac{\partial}{\partial \Pi_{jt}} \log\left(\frac{p_{lt}}{p_{kt}}\right) = \alpha_{lj} - \alpha_{kj}, \tag{4}$$

so we can recover from the logit coefficients the proportional impact on the relative odds of any two choices.

The choice model can be interpreted in either of two ways. The first interpretation is based on maximizing behavior, in the presence of random shocks to utility from different actions. This is similar to the Harsanyi (1973) approach to the purification of mixed strategies, employed recently by Fudenberg and Kreps (1993) in modeling the way players may learn to play mixed strategies. The utility of the player from action a_l at stage t is given by

$$U_{lt} = \sum_j \alpha_{lj} \Pi_{jt} + \epsilon_{lt}$$

where ϵ_{lt} follow an i.i.d. logistic distribution, and the player selects whichever alternative generates the highest utility. In this interpretation, the coefficient α_{lj} may be viewed as the way that a change in Π_{jt} affects the predictable component of utility from alternative l at t. For instance, this may be the outcome of "naive" Bayesian learning where players are attempting to update priors on perceived distributions resulting from different choices: Π_{jt} communicates information regarding the payoff the player may expect from action a_l a stage t. As it turns out, cross-coefficients frequently turn out to be significant, so it is important to be able to interpret them suitably. In the "signal processing" interpretation of these coefficients (within the utility-based framework), it is nevertheless natural to expect that the effect of a *ceteris paribus* increase in perceived payoff from action a_j on its own utility U_{jt} will be positive and will dominate the effect on the utility of other alternatives $U_{lt}, l \neq j$. This gives rise to two sets of restrictions on the logit coefficients:

(U1) $\alpha_{jj} > 0$, all $j = 1, \ldots, n-1$,

(U2) $\alpha_{jj} \geq \alpha_{lj}, l \neq j$.

Violation of (U1) would run contrary to any reasonable utility-based interpretation of the model, while (U2) is a secondary requirement which enhances the plausibility of the interpretation. A third conceivable restriction pertaining to the dominance of own- over cross-effects in signal processing is that own-perceived payoff plays a more important role in determining the utility of an action than that of any other alternative:

(U3) $\alpha_{jj} \geq \alpha_{jl}, l \neq j$.

The alternative interpretation of the choice model (1) is behavioral, in the spirit of psychological learning rules proposed by Bush and Mosteller (1955) or Suppes and Atkinson (1960). In such an approach, the primitive is the way that players allocate probabilities to different actions depending on experience from past plays. In such an approach it is more natural to express basic axioms in terms of the effects of changes in perceived payoffs on *absolute* rather than on *relative* probabilities of choice. Model (1) implies that

$$\eta_{lj} \equiv \frac{\partial p_{lt}}{\partial \Pi_{jt}} = p_{lt}\left(\alpha_{lj} - \sum_{k} \alpha_{kj} p_{kt} \right). \tag{5}$$

It is natural to expect in this interpretation that an increase in perceived payoff from action a_j will increase the probability assigned to this action, at the expense of the weight on the set of all other actions. Nevertheless, it is not necessary that the weight assigned to every other action decrease in this event. Other actions viewed as "similar" to a_j may also enjoy an increase in allocated weight, at the expense of all actions perceived as "dissimilar" to a_j.[3] It is possible that the players uses a multistage or

[3]For instance, it may be possible to partition the actions of an investor into a number of actions which involve investing some fraction of his portfolio in shares, and others where the entire portfolio is concentrated on fixed income securities. Then an increase in income from a portfolio which includes shares may cause the investor to assign greater weight to all portfolio choices that include shares.

nested procedure to allocate probability weights: having constructed a hierarchical sequence of partitions of the set of options available, allocate probability weight between the elements of the coarsest partition at the first stage, then allocate first-stage allocated weights within each element of this partition among elements of the subpartition, and so on. At each stage, the player will be discriminating among different elements of the current partition according to the relative values of the perceived payoffs from these (aggregated over constituent members in a suitable manner). Hence an increase in the perceived payoff from one action will tend to raise the weight allocated to other actions in the same element of the partition, and lower that for actions in other elements. Nevertheless, in any such procedure, it must be the case that the own-effects on absolute probabilities will be positive:

(B1) $\eta_{jj} > 0$, for all j.[4]

Cross-effects could either be positive or negative (depending on whether the actions belong to the same elements of the partition), but they ought always to be dominated by the own-effects:[5]

(B2) $\eta_{jj} \geq \eta_{lj}$, $l \neq j$

(B3) $\eta_{jj} \geq \eta_{jl}$, $l \neq j$.

In our empirical analysis, we will evaluate any given model according to either the utility-based (U1−U3) or the behavioral interpretation (B1−B3). Specifically, the plausibility of any of these interpretations will depend on which of these conditions is violated. Clearly, the positivity of own effects (U1 or B1) is quite fundamental to the plausibility of the interpretation, while the dominant diagonal conditions (U2, U3 or B2, B3) provide supplementary information.

When checking (U1−U3), it is possible to use information about the statistical significance of the estimated logit coefficients α_{lj}. For instance, insignificant coefficients may be ignored or believed to be set equal to zero. This will enable one to avoid rejecting models based on large sampling fluctuations. Unfortunately, however, computation of standard errors of the absolute probability effects η_{lj} is difficult, so checks on the behavioral interpretation will be based only on the sizes of estimated coefficients, without correcting for their statistical significance.

Finally, we discuss the definition of the perceived payoff variables Π_{lt}. Indeed, different theories of learning are distinguished primarily by this

[4] Note that this is implied by (U2).

[5] Conditions (B2) and (U2) are related: for instance, (U2) and $p_{jt} \geq p_{lt}$ implies $\eta_{jj} \geq \eta_{lj}$. But neither by itself implies the other, suggesting ways in which the utility-based and behavioral interpretations may conflict.

dimension of the model. Following Selten (1991), we may distinguish broadly between two classes of hypotheses:

Stimulus Learning. Here Π_{lt} is some average measure of payoffs experienced from play of action a_l in the past. Three specific models are considered for achieved earnings (1) time averages (TAAE), (2) moving averages (MAAE), and (3) vindication (VAE). In the empirical analysis, we consider simple time averages for achieved earnings in (1), and five period moving averages for (2), with the convention that the average payoff for any action not selected at all is set equal to zero. The VAE model looks back only one period: for any action it occupies the value 1 if it was selected and resulted in a win, $1/(n-1)$ if some other action was selected and resulted in a loss, and 0 otherwise (where n denotes the number of actions available). These models therefore differ primarily in the extent to which different past periods are weighted relative to one another.

Belief Learning. Here Π_{lt} is the expected payoff from action a_l at stage t, computed on the basis of assigned probabilities for the opponent's actions which are based on empirical frequencies of the latter's choices in past plays. Again, three specific models are considered: expected payoffs based on empirical frequencies which are (4) time averages (TAEP), (5) five period moving averages (MAEP), or (6) based on the previous period's play, the Cournot hypothesis (CEP).

Note that various models considered in the literature are special cases of the above, such as fictitious play (this corresponds to TAEP with negligible error variances), Bush–Mosteller models (VAE), Cournot adjustment (CEP with negligible error variances), or naive Bayesian learning (TAEP or MAEP depending on the nature of the priors).

4. EMPIRICAL RESULTS

4.1. *Choice Frequencies*

Table I presents overall frequencies of play of different choices in the four games. Choice frequencies for both row and column players are presented, along with those predicted by the minmax theory. Substantial deviations from the minmax predictions are apparent for all games, except perhaps for column players in Game 4. This is in contrast to the O'Neill (1987) experiment, where there appeared to be a close conformity between predicted and observed frequencies. Brown and Rosenthal (1990) pointed out in the context of that experiment, however, that chi-squared tests for equality of observed frequencies with the minmax predictions were deci-

TABLE I
Choice Frequencies (in Percentages)

Player	Choice 1	Choice 2	Choice 3	Choice 4	Choice 5	Choice 6
			Game 1			
Row	28.75	20.50	44.00	6.75		
Column	42.25	29.25	25.25	3.25		
(Minmax)	(37.25)	(25.00)	(37.25)	(0.00)		
			Game 2			
Row	42.50	24.00	4.50	8.25	17.75	3.00
Column	39.75	30.00	7.25	16.50	5.75	0.75
(Minmax)	(37.25)	(25.00)	(12.50)	(12.50)	(12.50)	(0.00)
			Game 3			
Row	34.75	13.25	42.25	9.75		
Column	34.25	32.35	28.75	4.75		
(Minmax)	(37.25)	(25.00)	(37.25)	(0.00)		
			Game 4			
Row	39.50	24.25	5.00	5.50	21.25	4.50
Column	33.75	23.75	12.50	16.75	10.25	3.00
(Minmax)	(37.25)	(25.00)	(12.50)	(12.50)	(12.50)	(0.00)

sively rejected at virtually every significance level, and we also obtain a similar result.[6] The rejection of the minmax theory at the aggregate level, and not just at the level of individual decision making, implies that it cannot even be treated as an acceptable *as if* theory.

A striking feature of the divergences (relative to minmax) is their qualitative similarity across games with changing payoffs. In games 1 and 3, the row player underplays choices 1 and 2 and overplays choice 3; the column player overplays choice 2 and underplays choice 3. In games 2 and 4, the row player underplays choices 3 and 4, while overplaying choice 5: collectively these three choices are underplayed. The column player overplays choice 4 and underplays choice 5. The fact that the more significant biases were replicated for strategically equivalent games across two different experiments (with different subjects) suggests that these were systematic rather than accidental. The *degree* of the biases of course vary

[6] This holds even if we neglect the fact that occasional plays of the dominated strategy take place, which tends to impart a large value to the chi-squared statistic. The only exception are the column players in Game 4, for whom minmax is nevertheless rejected at the 5% level.

across games with differing payoffs: chi-squared tests indicated a significant difference at the 5% level for the 4×4 game and for column players in the 6×6 game.

One interesting result is that the frequency of play of dominated actions was always higher in the games with the higher payoffs. To the extent that these plays are understood as the result of "mistakes" or "lapses of attention" within an otherwise optimizing framework (as in "decision cost" models), one would have expected the opposite result.

Across games with different number of choices available, the biases tend to differ significantly both qualitatively and quantitatively. For instance, the row players' inclination to underplay choice 2 in the 4×4 game disappears in the 6×6 game; the overplaying of choice 3 in the former is replaced by the underplaying of the choices 3–5 collectively in the latter. The first choice tends to be underplayed in the 4×4 games and overplayed in the 6×6 games. The hypothesis of equality of frequencies between these games is rejected at the 1% level for all except column players in the low payoff games (even for them it is rejected at the 5% level). Combined with the results of the preceding paragraph, this suggests that the structure of the game has a more marked effect on the nature of play than the scale of payoffs.

Our results therefore confirm the finding of Rapoport and Boebel (1992) that substantial divergences of observed frequencies of different choices from minmax predictions are found in constant sum games of moderate size (e.g., where each player has at least four actions to choose from). However, the specific frequencies observed in our experiment differ from theirs significantly, despite the similarity in the nature of the game played. This could be due to the different experimental settings employed (as discussed in the previous section) or to the fact that small changes in the structure of the game (e.g., with respect to the exact number of actions available) can have a large effect on observed frequencies. Similar to Rapoport and Boebel, we also find that the minmax theory fares better than alternative theories such as the "win-weighted" (WW) or "equiprobable" (EP) hypotheses.[7] The former postulates that players assign probability weights to different choices in proportion to the number of opportunities they present for winning, whilst the latter asserts that each action is assigned equal weight.

It will turn out from the analysis of the dynamics below that the minmax theory will be rejected also on the grounds that it predicts i.i.d. play, as in Brown and Rosenthal (1990), Mookherjee and Sopher (1994), and Rapoport and Boebel (1992). Nevertheless, it is possible that while the minmax

[7] This is indicated by the outcome of Wilcoxon signed-ranks tests for equality of distributions, using the same distance measures as in Rapoport–Boebel.

hypothesis is not an adequate explanation of the mode of behavior in constant sum games, it may nevertheless be an adequate predictor of empirical choice frequencies *in the long run*. Recent theoretical literature on learning (Fudenberg and Kreps, 1993; Jordan, 1991, 1993) has examined learning rules in which empirical frequencies *converge* to minmax predictions. Is there any evidence of such convergence in our experiment? Figure 1 presents the evolution of the time average of choice frequencies in different games, and Fig. 2 that of five-period moving averages of these frequencies. *It is evident from these that the time averages generally exhibit no tendency to converge to minmax, with the exception of the dominated action choice whose frequency appears to be vanishing as the game progresses*. *In most games, the divergences described above were established at early stages, and persistently maintained thereafter*. In some cases, the moving averages indicate some tendency for the divergences to narrow towards the very end, for instance for row players in games 1 and 3—apart from these exceptions the moving averages also reveal no tendency to converge to minmax predictions. It is of course conceivable that the number of stages played (40) was too short for patterns of convergence to be established. However, in line with Roth and Erev (1995) we tend to agree that the intermediate rather than long run properties of learning models are likely to be more relevant in explaining the pattern of play in our experiment. This motivates our attempt in the following section to discriminate between different learning hypotheses.

Note also the pattern of choices of the dominated action: they tend to be played to some extent within the first 10 stages, followed generally by a frequency that dies down thereafter. In some cases (such as row players in Games 1 and 3 or column players in Game 4) they are tried again around the middle to third-quarter stage, and discontinued thereafter. Specifically, their use tends to be bunched together around certain stages, thus casting doubt on explanations based on random preference shocks or mistakes in an optimizing framework.

4.2. *Dynamic Analysis*

Figures 1 and 2 have already illustrated the evolution of choice frequencies in different games. Before commencing to the logit estimates, it is useful to also look at the evolution of earnings achieved by players in past plays from different actions (the main ingredient for stimulus learning models), and of expected payoffs from different actions which are computed from predicting the opponent's play by past empirical frequencies (the ingredient of belief learning hypotheses). These are presented in Figs. 3 and 4. Both achieved earnings and expected payoff plots reveal substantial deviations from expected earnings under minmax play. Whereas min-

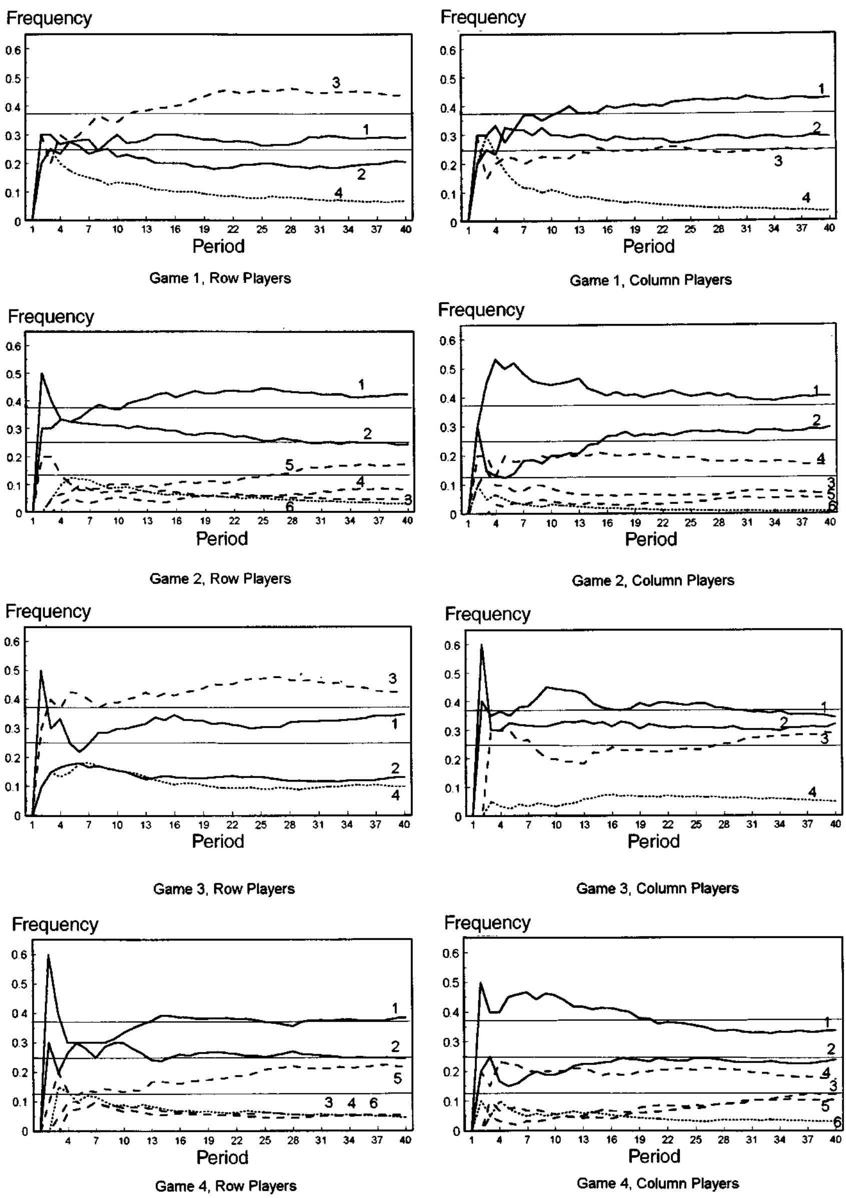

FIG. 1. Choice frequencies.

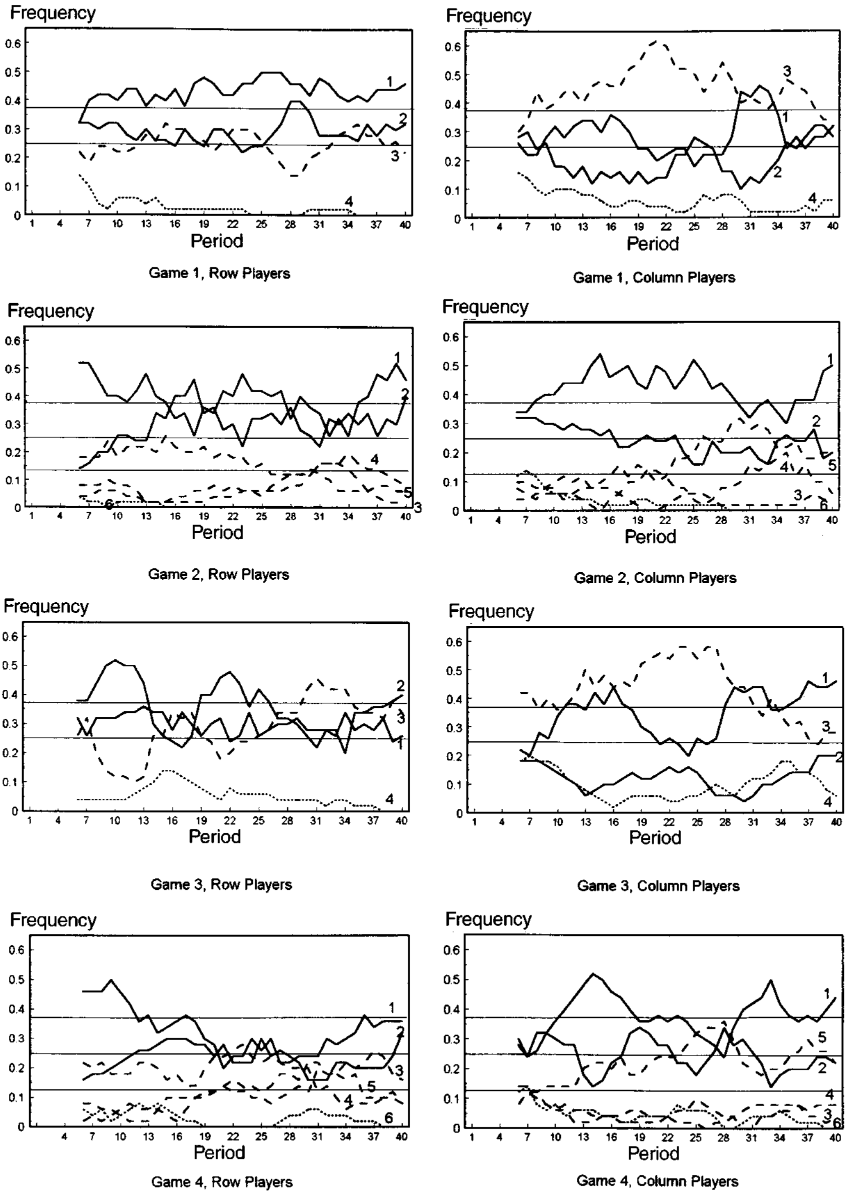

FIG. 2. Moving average choice frequencies.

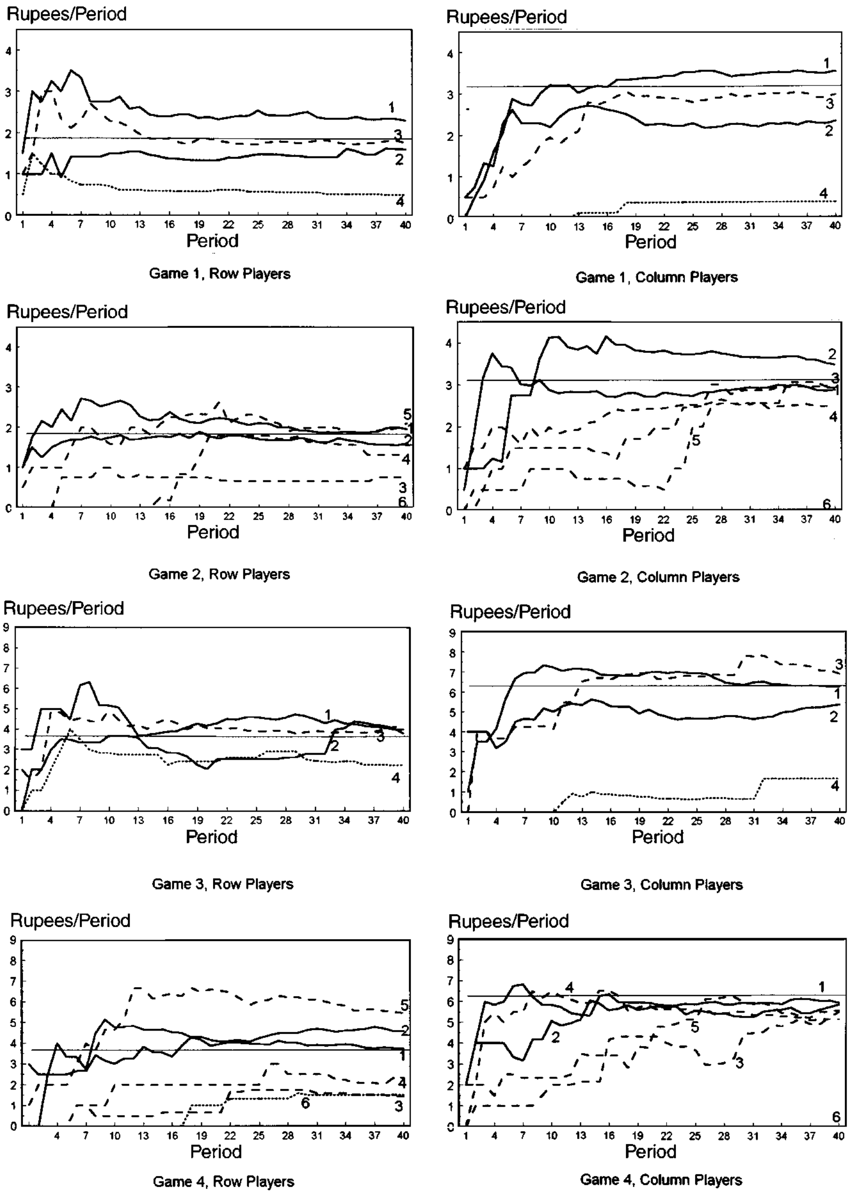

FIG. 3. Achieved earnings.

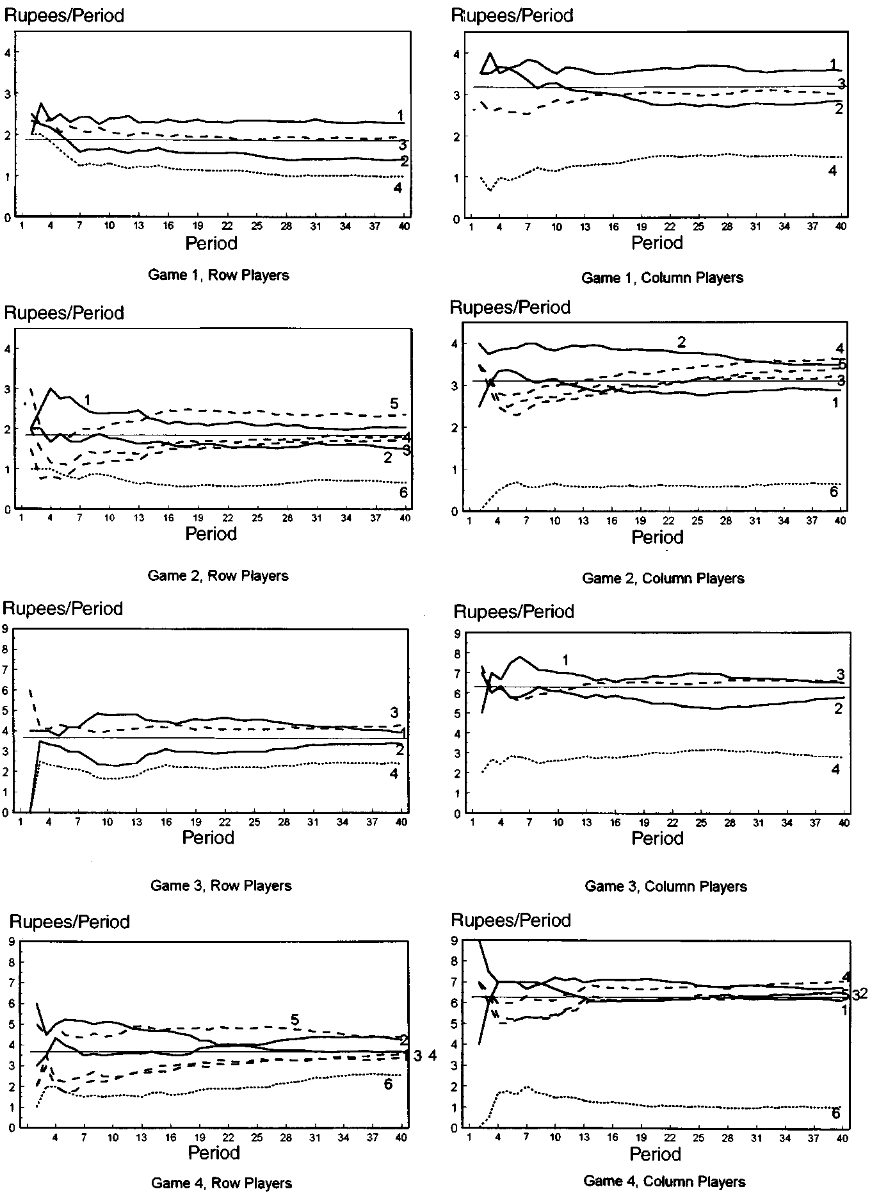

FIG. 4. Expected payoffs.

max predicts that all actions (except the dominated strategy) will be associated with the same expected payoff, both plots indicate substantial divergences in payoffs from different choices. Particularly interesting is that the pattern of many of these divergences correlates roughly with the divergences of observed choice frequencies from minmax predictions. For instance, choice 3, which is overplayed by row players in Games 1 and 3, is associated with higher earnings than under minmax. On the other hand, choice 2, which is underplayed, is associated with earnings that are low compared to those achieved from choice 3, or compared even to minmax earnings. With few exceptions, a similar correspondence is observed for the choice frequencies which are particularly divergent from minmax predictions (such as the relative biases between choices 3, 4, and 5 in the 6×6 games). Similar patterns are observed in the expected payoff plots as well, except that the correspondence is less marked, partly because they exhibit markedly smaller degrees of variability than the achieved earnings plots. Overall, these plots suggest that the pattern of play may be related to experience acquired by the players. To investigate this more thoroughly, we turn to the estimated logit equations.

4.2.1. *Model Selection*. Section 3 described six different hypotheses regarding the learning pattern that could potentially be tested from the data for different players and for different games. A number of other choices have to be made with respect to the exact specification of the logit equations. To keep the exposition within a manageable limit, it is necessary to restrict the alternative models to be discussed in detail. We therefore address issues involved in model selection that motivate the particular versions discussed in detail in the following section.

We start by trying to restrict alternative variants of the belief learning and stimulus learning hypotheses. Consider for instance column players in Game 1. Table II presents multinomial logit results corresponding to the six different models identified in Section 3. The table reports values of the logit coefficients, as well as n, the number of observations; log L, the maximized log-likelihood; AIC, the value according to the *A*kaike *i*nformation *c*riterion used for discriminating among models; and PR-sq, the value of a pseudo-R^2 measure of goodness of fit.[8] Player specific dummies were not included since they appeared to be insignificant and added little additional explanatory power. The reported results are for multinomial logits estimated without constant terms. We also estimated all models with

[8]The pseudo-R^2 measure used is the proportionate increase in the log likelihood by addition of the independent variables, over and above inclusion of the constant terms. For models estimated without a constant term, the pseudo R^2 measure is the proportionate increase in the log likelihood over the likelihood for a model in which all choices are made with equal probability. We report results for the second (no-constant) type of logit.

TABLE II
Estimated Logits Column Players, Game 1

	Time Average Achieved Earnings	Moving Average Achieved Earnings	Vindication	Time Average Expected Payoffs	Moving Average Expected Payoffs	Cournot
n	390	350	390	390	350	390
log L	−415.66	−400.78	−453.70	−439.56	−391.88	−447.87
AIC	2.18	2.34	2.37	2.30	2.29	2.34
PR-sq	0.23	0.17	0.16	0.19	0.19	0.17
α_{11}	1.23*	0.83*	3.10*	0.15	0.29	0.57*
α_{12}	0.27	0.18	2.83*	1.96^{+}	0.66	1.54*
α_{13}	−0.26	0.10	2.45*	1.34*	1.20	0.63*
α_{14}	−0.45	3.84*	2.25*	−2.66^{+}	−0.55	−1.88*
α_{21}	0.65*	0.71*	3.00*	−0.17*	0.11	0.45*
α_{22}	0.77^{+}	0.25	2.00^{+}	1.55	−0.10	1.32^{+}
α_{23}	−0.08	0.11	1.94*	0.92^{+}	0.92^{+}	0.37^{+}
α_{24}	−0.45	3.87*	2.04	−1.56	0.75	−1.42?
α_{31}	0.56^{+}	0.60*	2.60*	−0.32	0.14	0.45*
α_{32}	0.57?	0.25	2.37*	1.63?	0.09	1.43^{+}
α_{33}	0.15	0.22	1.89*	1.06^{+}	0.90^{+}	0.45*
α_{34}	−0.51	3.90*	1.55	−1.73	0.42	−1.71^{+}

Note. *denotes significance at the 1% level, $^{+}$ at the 5% level, and ? at the 10% level.

constant terms, but certain variables (mainly payoff variables for the dominated strategy in the expected payoff logits) turned out to be highly collinear with the constant term. There is no qualitative difference between the two types of estimates (constant included or not included) in terms of the relative performance of the different models.

It is evident that the time average version of the achieved earnings model performs better than the moving average or vindication versions. This is indicated by both statistical measures of goodness of fit, and the criterion of signs and significances of estimated coefficients. Note that properties U1 and U2 are satisfied by all three equations, and U3 is satisfied by the first and second equations in the time average model. Moreover, the equations explaining choices 1 and 2 are characterized by significant own-effects (at the 5% level). In the moving average model the cross-effects of varying the payoff of the dominated action on the other actions are significant, but have the wrong signs. Earnings for choice 1 are also significant in all three equations. Many coefficients in the vindication model turn out to be statistically significant, and U1 is satisfied in all three equations, but conditions U2 and U3 are violated for the second and third choice. The coefficients on the dominated strategy are large, and significant in Eq. (1), but of the wrong sign, as in the moving average model.

With respect to the three versions of the expected payoff model, the Cournot model fares the poorest from the standpoint of goodness of fit. Though many of the estimated coefficients are highly significant and condition U1 is satisfied throughout, conditions U2 and U3 are routinely violated. Of the remaining two versions, the moving average model achieves the same fit (judged by the pseudo-R^2), and a slightly better (lower) value of the Akaike criterion. The pattern of estimated coefficients for either version does not lend much support to either hypothesis.

The results of Table II thus indicate that the only reasonable model to explain the behavior of column players in Game 1 is the one based on time average achieved earnings. It succeeds in explaining the tendency for these players to overplay their first two choices (though not the underplaying of the third choice).

In similar vein, it turns out that for both categories of players, and for all games, attention can be restricted to the time average versions of either the belief or stimulus learning hypotheses: other versions are dominated by at least one of these. In view of this, we hereafter present only the time average versions of the achieved earnings (denoted TAAE) and expected payoff (TAEP) models.

Another issue is the inclusion of cross-effects. Do variations in payoffs from one action affect choices of other actions? The McKelvey–Palfrey (1993) formulation of discrete choice behavior, for instance, assumes that these cross-effects do not arise. In the interests of parsimony, it makes sense to exclude them from the estimated equations if they turn out to be insignificant and ignore the *a priori* restrictions concerning cross-effects while discriminating between different theories. Table III reports the results of testing for the restriction that all cross-effects are absent in either TAAE and TAEP logit equations, using a standard likelihood ratio test, as well as the Akaike criterion. With a few exceptions (such as row players in Game 4), the restriction is rejected at the 5% level. The Akaike Information Criterion also points to the unrestricted version in most cases (by "rejection" in the case of the AIC, we simply mean that the unrestricted model has a lower AIC value). In what follows, we therefore report results for the unrestricted versions of these two models and examine possible violations of the *a priori* restrictions (U2, U3, B2, B3) associated with different hypotheses.

As mentioned above, collinearity between the payoff from the dominated action and the constant term often led to a choice between inclusion of one of these at the expense of the other. It turns out that for both TAAE and TAEP models, dropping the constant term and retaining the payoff from the dominated action provided coefficient estimates which

TABLE III
Testing Zero Cross-Effect Restrictions

Player Type	Game	Model	LR Test 1%	LR Test 5%	Akaike Criterion
Row	1	TAAE	R	R	R
Row	1	TAEP	R	R	R
Row	2	TAAE	R	R	R
Row	2	TAEP	R	R	R
Row	3	TAAE	R	R	R
Row	3	TAEP	NR	NR	NR
Row	4	TAAE	NR	NR	NR
Row	4	TAEP	NR	NR	NR
Column	1	TAAE	R	R	R
Column	1	TAEP	R	R	R
Column	2	TAAE	NR	R	NR
Column	2	TAEP	NR	R	NR
Column	3	TAAE	R	R	R
Column	3	TAEP	R	R	R
Column	4	TAAE	NR	R	NR
Column	1	TAEP	R	R	R

Note. R and NR denote the hypothesis of zero cross effects either rejected or not rejected, respectively.

were more plausible (in terms of criteria U1−U3 and B1−B3) and statistically significant, with no appreciable change in the Akaike criterion. In what follows, therefore, as in Table II, we report the results of these no-constant term versions of the TAAE and TAEP models.

4.2.2. *Discussion of Results*. Tables IV−IX present estimated logit equations for the TAAE and TAEP models for different categories of players and games.[9] Overall, the TAAE model provides consistently reasonable estimates, as well as a statistically significant explanation of the pattern of play, for both row and column players, and in all games. On the other hand, this is true for the TAEP model only in the case of row players in Game 1: see Table X which summarizes the number of violations of conditions U1−U3 and B1−B3 for either model. Conditions U1, B1 and U2 are never violated by the TAAE model, while the remaining ‘‘dominant diagonal’’ conditions B2 and B3 for the absolute probability effects are violated relatively infrequently (for at most one equation, with the exception of row players in Game 4). Violation of the relative odds condition U3 is however more frequent, especially in the 6×6 games.

[9]As mentioned previously, player-specific dummies for any particular category of player and for a given game were collectively insignificant, and were thus not included. Moreover, the reported logit equations exclude the constant term, and occasionally the variable representing payoff from the dominated action owing to problems of collinearity.

TABLE IV
TAAE and TAEP Estimates for Row Players in 4 × 4 Games

	Payoff from Choice 1 $\alpha.1$	Payoff from Choice 2 $\alpha.2$	Payoff from Choice 3 $\alpha.3$	Payoff from Choice 4 $\alpha.4$	Payoff from Choice 1 $\eta.1$	Payoff from Choice 2 $\eta.2$	Payoff from Choice 3 $\eta.3$	Payoff from Choice 4 $\eta.4$
Game 1								
TAAE Equation: $n = 390$, LL $= -441.80$, $PR^2 = 0.183$, AIC $= 2.312$								
Estimated Effect on Probability of Game 1 Choice No.:								
1	0.61*	−0.25?	0.49+	−0.42*	4.39	−2.80	−2.10	−3.45
2	0.56*	−0.05	0.17	−0.36+	2.10	2.10	−8.05	−1.23
3	0.38+	−0.16	0.88*	−0.24?	−3.40	−0.33	13.95	2.65
4					−3.09	1.03	−3.80	2.03
TAEP Equation: $n = 390$, LL $= -438.91$, $PR^2 = 0.188$, AIC $= 2.297$								
Estimated Effect on Probability of Game 1 Choice No.:								
1	1.18*	1.21	0.22	−2.64+	5.89	−13.11	−1.21	9.57
2	1.17*	3.49*	−0.36	−5.09*	4.00	37.79	−12.76	−43.40
3	0.90*	1.37	0.62*	−2.66+	−3.30	−13.03	15.74	13.77
4					−6.58	−11.25	−1.77	20.07
Game 3								
TAAE Equation: $n = 390$, LL $= -464.05$, $PR^2 = 0.14$, AIC $= 2.426$								
Estimated Effect on Probability of Game 3 Choice No.:								
1	−0.02	0.14+	0.31*	−0.05	0.74	−0.17	1.45	−0.40
2	−0.10	0.28*	0.00	0.00	−0.78	1.79	−3.55	0.51
3	0.05	0.14	0.38*	−0.05	−0.37	−0.21	4.72	−0.49
4					0.40	−1.41	−2.62	0.38
TAEP Equation: $n = 390$, LL $= -465.57$, $PR^2 = 0.14$, AIC $= 2.434$								
Estimated Effect on Probability of Game 3 Choice No.:								
1	0.02	−0.32	0.41*	0.24	2.36	−4.17	0.27	3.37
2	−0.16	−0.51	0.11	0.96	−1.48	−4.11	−3.87	10.83
3	−0.08	−0.05	0.58*	−0.16	−1.35	6.33	7.52	−12.80
4					0.47	1.95	−3.92	−1.39

Note. LL denotes the log-likelihood, PR^2 denotes the pseudo-R^2, and AIC denotes the value under Akaike information criterion. *denotes significance at the 1% level, + at the 5% level, and ? at the 10% level for α estimates. Significances for η variables were not available. − denotes variables that were dropped due to collinearity.

In contrast, the TAEP model generates negative own-effects quite often: in every 6 × 6 game, conditions U1 and B1 are violated at least once, while all other dominant diagonal conditions are violated for a majority of choices. In other words, the model predicts that a higher expected payoff from a given action often results in a significant lowering of weight on that

TABLE V
TAAE and TAEP Estimates for Column Players in 4 × 4 Games

	Payoff from Choice 1 $\alpha.1$	Payoff from Choice 2 $\alpha.2$	Payoff from Choice 3 $\alpha.3$	Payoff from Choice 4 $\alpha.4$	Payoff from Choice 1 $\eta.1$	Payoff from Choice 2 $\eta.2$	Payoff from Choice 3 $\eta.3$	Payoff from Choice 4 $\eta.4$
Game 1								
TAAE Equation: $n = 390$, LL $= -415.66$, PR$^2 = 0.23$, AIC $= 2.178$								
Estimated Effect on Probability of Game 1 Choice No.:								
1	1.23	0.27	−0.26	−0.45	16.13	−9.01	−6.96	0.02
2	0.64*	0.77$^+$	−0.08	−0.45	−6.09	8.39	0.45	0.02
3	0.56$^+$	0.57?	0.15	−0.51	−7.28	2.19	6.20	−1.50
4					−2.76	−1.57	0.31	1.46
TAEP Equation: $n = 390$, LL $= -439.56$, PR$^2 = 0.19$, AIC $= 2.300$								
Estimated Effect on Probability of Game 1 Choice No.:								
1	0.15	1.96$^+$	1.34*	−2.66$^+$	9.07	11.16	10.02	−27.17
2	−0.17	1.55	0.92$^+$	−1.56	−3.08	−3.98	−5.35	13.37
3	−0.32	1.63?	1.06$^+$	−1.73	−6.20	−1.67	−1.08	7.25
4					0.21	−5.51	−3.58	6.56
Game 3								
TAAE Equation: $n = 390$, LL $= -455.94$, PR$^2 = 0.16$, AIC $= 2.384$								
Estimated Effect on Probability of Game 3 Choice No.:								
1	−0.15?	0.14	0.16$^+$	−0.42*	4.26	−1.23	−2.31	−2.33
2	−0.02	0.21?	0.25*	−0.28*	−1.15	1.09	0.73	2.32
3	0.09	0.20?	0.32*	−0.41*	−3.04	0.98	2.66	−1.67
4					−0.07	−0.84	−1.08	1.67
TAEP Equation: $n = 390$, LL $= -466.11$, PR$^2 = 0.14$, AIC $= 2.436$								
Estimated Effect on Probability of Game 3 Choice No.:								
1	0.78*	2.24*	−0.29	−3.03*	−1.17	−5.22	0.27	9.60
2	0.84*	2.32*	−0.25	−3.24	0.84	−2.33	1.54	2.27
3	0.96*	3.05*	−0.41$^+$	−4.27*	4.20	18.91	−3.23	−27.59
4					−3.87	−11.36	1.41	15.72

Note. For explanations, see the notes to Table IV.

action, either in absolute terms or relative to some other action. In terms of statistical fit or nonnested-model-selection criteria such as the Akaike criterion, the same result emerges: the TAAE model provides a better explanation of the data, except only for row players in Game 1.

The same inference can be made from the significant frequency of plays of the dominated action. The plots in Figs. 1 and 2 give the impression that these were tried at the earlier stages of the game, and the players "learned" from their experience with this choice to shift weight to the

TABLE VI
TAAE and TAEP Estimates for Row Players, Game 2

TAAE Equation: $n = 390$, LL $= -543.84$, $PR^2 = 0.22$, AIC $= 2.892$

	Payoff from Choice 1	Payoff from Choice 2	Payoff from Choice 3	Payoff from Choice 4	Payoff from Choice 5	Payoff from Choice 6
Estimated Effect on Relative Likelihood of Choice No.:						
	$\alpha.1$	$\alpha.2$	$\alpha.3$	$\alpha.4$	$\alpha.5$	$\alpha.6$
1	0.77*	0.01	0.00	1.08	0.09	—
2	0.27	0.30?	0.08	0.88	0.17	—
3	−0.03	−0.09	1.00*	0.95	-0.62^{+}	—
4	−0.08	0.16	0.34	1.19?	0.01	—
5	0.17	−0.01	0.14	1.18?	0.29^{+}	—
Estimated Effect on Absolute Likelihood of Choice No.:						
	$\eta.1$	$\eta.2$	$\eta.3$	$\eta.4$	$\eta.5$	$\eta.6$
1	15.12	−3.13	−4.98	2.53	−0.57	—
2	−3.46	5.19	−0.89	−3.37	1.60	—
3	−2.00	−0.78	3.97	−0.32	−3.26	—
4	−4.08	0.63	1.84	1.40	−0.77	—
5	−4.34	−1.66	0.41	2.83	3.31	—
6	−1.24	−0.25	−0.35	−3.06	−0.31	—

TAEP Equation: $n = 390$, LL $= -526.43$, $PR^2 = 0.25$, AIC $= 2.828$
Estimated Effect on Relative Likelihood of Choice No.:

	$\alpha.1$	$\alpha.2$	$\alpha.3$	$\alpha.4$	$\alpha.5$	$\alpha.6$
1	0.21	0.03	−1.19?	1.69*	1.19^{+}	-1.83^{+}
2	−0.1	1.02	−0.72	1.12?	0.58	−1.66
3	−0.03	0.29	−0.57	1.59	−0.06	−1.84
4	−0.61?	0.36	−0.14	1.05	0.55	−0.99
5	−0.68*	0.45	−1.14	1.20	1.71^{+}	−1.94?
Estimated Effect on Absolute Likelihood of Choice No.:						
	$\eta.1$	$\eta.2$	$\eta.3$	$\eta.4$	$\eta.5$	$\eta.6$
1	13.58	−14.88	−11.56	14.10	8.45	−6.16
2	−0.01	15.36	4.75	−5.72	−9.87	0.60
3	0.36	−0.41	1.57	1.04	−4.73	−0.70
4	−4.13	−0.17	6.42	−2.54	−3.64	5.73
5	−10.13	1.24	−3.94	−2.81	12.76	−4.53
6	0.33	−1.14	2.75	−4.07	−2.97	5.05

Note. For explanations, see notes to Table IV.

TABLE VII
TAAE and TAEP Estimates for Row Players in Game 4

TAAE Equation: n = 390, LL = −548.41, PR^2 = 0.22, AIC = 2.941

	Payoff from Choice 1	Payoff from Choice 2	Payoff from Choice 3	Payoff from Choice 4	Payoff from Choice 5	Payoff from Choice 6
	Estimated Effect on Relative Likelihood of Choice No.:					
	$\alpha.1$	$\alpha.2$	$\alpha.3$	$\alpha.4$	$\alpha.5$	$\alpha.6$
1	0.18^{+}	0.24*	0.59	0.11	0.07	−0.41
2	0.00	0.33*	0.63?	0.13?	0.04	−0.46^{+}
3	0.07	−0.03	0.70?	0.01	−0.12	−0.32
4	−0.13	0.04	0.65?	0.21^{+}	0.01	−0.46?
5	0.04	0.20^{+}	0.46	0.10	0.10	−0.26
	Estimated Effect on Absolute Likelihood of Choice No.:					
	$\eta.1$	$\eta.2$	$\eta.3$	$\eta.4$	$\eta.5$	$\eta.6$
1	4.11	0.87	1.41	0.07	0.67	−1.58
2	−1.84	2.72	1.84	0.53	−0.32	−2.18
3	−0.03	−1.24	0.73	−0.49	−0.87	0.25
4	−1.13	−0.98	0.53	0.56	−0.24	−0.49
5	−0.76	−0.38	−2.00	−0.18	1.00	2.34
6	−0.34	−0.98	−2.49	−0.49	−0.24	1.67

TAEP Equation: n = 390, LL = −557.27, PR^2 = 0.20, AIC = 2.986

	Estimated Effect on Relative Likelihood of Choice No.:					
	$\alpha.1$	$\alpha.2$	$\alpha.3$	$\alpha.4$	$\alpha.5$	$\alpha.6$
1	0.45*	0.53	0.80	−0.28	−0.61	−0.26
2	0.26	0.40	0.83	−0.51	−0.36	−0.12
3	0.30	−0.29	−0.13	0.09	−0.16	0.30
4	−0.06	0.82	0.62	0.53	−0.93	−1.13
5	0.15	0.34	0.79	−0.35	−0.24	−0.30
	Estimated Effect on Absolute Likelihood of Choice No.:					
	$\eta.1$	$\eta.2$	$\eta.3$	$\eta.4$	$\eta.5$	$\eta.6$
1	6.54	4.77	3.45	−0.20	−6.78	−0.68
2	−0.59	−0.22	2.84	−5.70	1.90	2.98
3	0.08	−3.50	−4.21	1.83	1.39	2.71
4	−1.89	2.26	−0.51	4.43	−2.70	−4.88
5	−2.86	−1.47	1.64	−1.59	4.22	−1.22
6	−1.28	−1.84	−3.21	1.24	1.97	1.09

Note. For explanations, see notes to Table IV.

TABLE VIII
TAAE and TAEP Estimates for Column Players, Game 2

.	Payoff from Choice 1	Payoff from Choice 2	Payoff from Choice 3	Payoff from Choice 4	Payoff from Choice 5	Payoff from Choice 6
TAAE Equation: $n = 390$, LL $= -502.42$, PR$^2 = 0.28$, AIC $= 2.679$						
Estimated Effect on Relative Likelihood of Choice No.:						
	$\alpha.1$	$\alpha.2$	$\alpha.3$	$\alpha.4$	$\alpha.5$	$\alpha.6$
1	1.29$^+$	9.18	−0.51	−0.02	2.39	—
2	1.09?	9.28*	−0.54	0.01	2.39	—
3	0.64	8.63*	−0.09	0.17	2.46*	—
4	1.13$^+$	8.99*	−0.57	0.12	2.44	—
5	0.78	8.54*	−0.45	0.05	2.95*	—
Estimated Effect on Absolute Likelihood of Choice No.:						
	$\eta.1$	$\eta.2$	$\eta.3$	$\eta.4$	$\eta.5$	$\eta.6$
1	6.86	5.84	−0.75	−1.99	−1.10	—
2	−0.82	7.41	−1.46	−0.60	−0.83	—
3	−3.46	−2.92	2.91	1.01	0.31	—
4	0.21	−0.71	−1.30	1.48	0.37	—
5	−1.94	−2.84	0.24	0.11	3.06	—
6	−0.84	−6.77	0.37	−0.02	−1.81	—
TAEP Equation: $n = 390$, LL $= -526.58$, PR$^2 = 0.25$, AIC $= 2.83$						
Estimated Effect on Relative Likelihood of Choice No.:						
	$\alpha.1$	$\alpha.2$	$\alpha.3$	$\alpha.4$	$\alpha.5$	$\alpha.6$
1	0.94	85.94*	0.03	1.02	−0.17	−86.22*
2	0.95	86.89*	−0.21	1.48	−0.43	−87.49*
3	0.64	85.56*	−0.17	0.96	−0.58	−85.57*
4	0.49	86.32*	−0.44	1.25	0.04	−86.50*
5	0.41	88.54*	0.34	0.76	−1.10	−88.90*
Estimated Effect on Absolute Likelihood of Choice No.:						
	$\eta.1$	$\eta.2$	$\eta.3$	$\eta.4$	$\eta.5$	$\eta.6$
1	5.19	46.95	5.82	−5.92	4.98	−4.47
2	4.22	33.75	−2.81	9.33	−4.04	−41.47
3	1.23	−1.49	−0.39	−1.52	−2.06	3.90
4	−5.27	9.16	−5.34	1.34	5.53	−6.48
5	−2.30	15.96	2.62	−2.35	−4.63	−16.06
6	−0.61	−64.32	0.09	−0.88	0.22	64.58

Note. For explanations, see the notes to Table IV.

TABLE IX
TAAE and TAEP Estimates for Column Players, Game 4

.	Payoff from Choice 1	Payoff from Choice 2	Payoff from Choice 3	Payoff from Choice 4	Payoff from Choice 5	Payoff from Choice 6
TAAE Equation: $n = 390$, LL $= -579.29$, PR$^2 = 0.17$, AIC $= 3.07$						
Estimated Effect on Relative Likelihood of Choice No.:						
	$\alpha.1$	$\alpha.2$	$\alpha.3$	$\alpha.4$	$\alpha.5$	$\alpha.6$
1	0.35*	0.08	0.08	−0.01	−0.07	—
2	0.12	0.27*	0.06	0.00	−0.07	—
3	0.05	0.12	0.24*	−0.10	−0.05	—
4	0.04	0.12	0.02	0.20?	−0.07	—
5	−0.10	0.19^{+}	0.09	0.03	0.06	—
Estimated Effect on Absolute Likelihood of Choice No.:						
	$\eta.1$	$\eta.2$	$\eta.3$	$\eta.4$	$\eta.5$	$\eta.6$
1	6.77	−2.22	−0.13	−1.04	−1.03	—
2	−0.70	2.95	−0.57	−0.49	−0.72	—
3	−1.24	−0.32	1.95	−1.51	1.12	—
4	−1.83	−0.43	−1.07	3.00	−0.51	—
5	−2.56	0.45	0.06	0.10	1.02	—
6	−0.45	−0.44	−0.25	−0.06	0.12	—
TAEP Equation: $n = 390$, LL $= -596.70$, PR$^2 = 0.15$, AIC $= 3.19$						
Estimated Effect on Relative Likelihood of Choice No.:						
	$\alpha.1$	$\alpha.2$	$\alpha.3$	$\alpha.4$	$\alpha.5$	$\alpha.6$
1	0.21	0.69	−0.17	0.09	0.03	−0.16
2	0.43^{+}	0.44	0.09	−0.03	−0.26	−0.18
3	0.38?	0.75	0.21	0.08	−0.58	−0.56
4	0.35?	0.75	−0.10	−0.12	0.01	−0.52
5	0.39?	0.62	0.64?	0.03	−0.88	−0.75
Estimated Effect on Absolute Likelihood of Choice No.:						
	$\eta.1$	$\eta.2$	$\eta.3$	$\eta.4$	$\eta.5$	$\eta.6$
1	−3.68	2.35	−7.06	2.49	9.19	5.76
2	2.63	−4.28	1.21	−1.10	−1.12	3.58
3	0.76	1.62	2.14	0.80	−4.59	−2.87
4	0.52	2.17	−2.33	−2.28	3.73	−3.17
5	0.73	0.00	6.16	0.14	−6.84	−4.30
6	−0.96	−1.86	−0.12	−0.05	0.64	0.99

Note. For explanations, see notes to Table IV.

TABLE X
Violation of a Priori Restrictions

	Number of Violations of:						
	U0#	U1	U2	U3	B1	B2	B3
(4 × 4 games)							
Row, Game 1, TAAE	1	0	0	1	0	1	0
Row, Game 1, TAEP	0	0	0	1	0	0	1
Row, Game 3, TAAE	1	0	0	1	0	1	1
Row, Game 3, TAEP	2	1	1	2	2	2	3
Col, Game 1, TAAE	1	0	0	1	0	0	0
Col, Game 1, TAEP	2	0	2	2	2	2	3
Col, Game 3, TAAE	0	0	0	0	0	1	0
Col, Game 3, TAEP	1	1	3	2	3	3	3
(6 × 6 games)							
Row, Game 2, TAAE	0	0	0	3	0	1	1
Row, Game 2, TAEP	4	1	2	2	1	3	2
Row, Game 4, TAAE	1	0	0	4	0	2	1
Row, Game 4, TAEP	4	2	3	5	2	3	3
Col, Game 2, TAAE	2	0	0	4	0	0	0
Col, Game 2, TAEP	3	2	4	4	2	3	4
Col, Game 4, TAAE	1	0	0	1	0	1	0
Col, Game 4, TAEP	5	2	5	4	4	4	4

Note. # U0 denotes the restriction that own effects are positive and significant at 10%. Violations occurring at second and higher decimal places are ignored.

other choices available. This is consistent with a stimulus learning model: indeed, Tables IV−VII attest to this, since the coefficient of the payoff from the dominated action on all other choices is uniformly negative. On the other hand, a belief learning model must interpret plays of the dominated action as a departure from ‘‘rationality’’ or as the result of random shocks to player’s utilities. The latter interpretation would lead one to expect occasional plays of the dominated action throughout the course of the game. Instead, their play tended to bunch together, generally at earlier stages of the game. Moreover, the plots indicate that the payoff loss from playing the dominated action was quite significant, so the sizes of these shocks or of the necessary departures from rationality would have to be quite large.

The TAAE model does succeed in explaining many of the prominent biases in observed choice frequencies noted in Section 4.1 in terms of payoffs experienced in the past from these choices, barring a few (such as the tendency for row players to underplay their second choice in Game 1, or overplay their fifth choice in Game 4). For row players the success of the TAAE model appears to be slightly better in the 6×6 games. In general, the statistical fit is somewhat higher in the 6×6 games. Thus there is a little evidence to suggest that increased strategic complexity causes players to increase reliance on stimulus learning, but this is tenuous at best. There is no discernible pattern with respect to changing payoff scale: indeed, it is surprising that coefficient estimates change so much from game to game. Moreover, they often change appreciably *within* the course of any given game.[10]

5. CONCLUSION

To summarize, the main findings of this paper are the following:

1. In constant sum games with at least four choices available to each player, there are substantial departures from minmax predictions, both in terms of observed choice frequencies, as well as in the mode of behavior: there is evidence that subjects learn from past payoff experience. This is in striking contrast to our earlier experiment with a constant sum game where each player has only two choices available.

2. The evidence is more consistent with models of stimulus rather than belief learning: i.e., there is more favorable evidence for the hypothesis that players allocate weight across different actions in accordance with the relative payoffs experienced in the past, compared to the hypothesis that they select a best response to some beliefs about opponent play based on past experience.

3. Players seem to exhibit long rather than short memories, in both versions of stimulus or belief learning: simple time averaging processes track the data better than do moving average processes or one-period reactions. If anything, play early on appears to exercise an important effect on observed choice frequencies.

[10] The plots suggest considerably greater variability in payoff experience and in choice frequencies at the earlier stages of the game. This suggests that the players learn ‘‘more’’ early on in the game, and the learning pattern is not stable between early and late stages of the game. This does turn out to be the case when the data is split between the first half of the game and the second half, and the TAAE equation is estimated on each set separately. We did not succeed in capturing this ‘‘early learning’’ effect by interacting the stage of the game with the payoffs experienced.

4. Learning parameters appear to be unstable both across similar games and within (early versus late period) any given game. Effects of varying the number of choices available (from four to six) or doubling the scale of payoffs do not yield any clear implications for the nature of learning.

It is nevertheless important to qualify the above results by mentioning that all our tests were for the joint hypothesis of a specific learning hypothesis and the multinomial logit formulation of choice behavior.

Our work is closely related to a number of recent papers exploring learning in experimental games. It is most closely related to Roth and Erev (1995), who find that a process very similar to our stimulus learning model is successful in tracking the dynamics of three different experiments.[11] Merlo and Schotter (1992) find that when subjects "learn while they earn" (as in our experiment), they do not end up with "optimal" decisions even at the end of 75 rounds; players behave adaptively with respect to their own payoff experience and ignore information regarding opponents' choices.[12] These are also consistent with our results, though one respect in which their results differ is that their players appear to have relatively short memories (incorporating the payoff experience of only the last two periods).

All of the above-mentioned games offered each player a large number of available actions to choose from. In contrast, McCabe *et al.* (1993) and Mookherjee and Sopher (1994) considered games where each player had two actions to choose from.[13] When subjects were provided with complete information regarding opponents choices and payoffs, the minmax hypothesis could not be rejected, either in terms of its predictions regarding observed frequencies or the behavior of subjects. This suggests that the number of choices available has an important effect on the nature of play: whenever each player has at least four choices, and minmax play involves frequencies that are not focal in any sense, players seem to use simple learning procedures based only on their own payoff experience. As noted in the Introduction, the divergence from minmax and the lack of serial

[11] In their model, players allocate weight according to the *cumulative* past payoffs from each action, rather than the respective *average* earnings. The three experiments pertained to similar extensive form games, with differing outcomes: in two convergence to a perfect Nash equilibrium was attained, unlike the third. Each game involved players selecting offers lying between 0 and 1000.

[12] They study a "tournament" game where twelve players each had to select from a hundred numbers each.

[13] The latter consider two-person repeated matching pennies, while the former study games with a larger number of players where the payoff to each player depends on her own choice, as well as those of a subgroup of remaining players.

independence in the pattern of play in 4×4 or 5×5 constant sum games was also the main result of Brown and Rosenthal (1990) and Rapoport and Boebel (1992).

Boylan and El-Gamal (1993) find evidence favoring the hypothesis of fictitious play over the Cournot one-period-lagged hypothesis in a diverse number of experiments. This is consistent with our result concerning the relative lengths of memory: the fictitious play model generally performed better than the Cournot hypothesis. We found however that the time averaged stimulus learning model in turn generally performed better than fictitious play. It would be interesting to examine whether this is true in a wider range of experiments.

Our approach based on quantal response learning generalizes the QRE notion of McKelvey and Palfrey (1993) to explore the extent to which divergences from equilibrium predictions can be explained by preference shocks or departures from full rationality. The main distinction pertains to the extent to which players can learn from previous experience in playing the game. The QRE notion presumes that players succeed in coordinating on quantal response strategies which are best replies to one another, which preclude the possibility of learning about appropriate quantal response strategies from past experience. The presence of serial dependence in observed patterns of play in games of at least moderate complexity suggest that such forms of learning need to be incorporated. Nevertheless, it is conceivable that quantal response learning procedures converge in the long run to a QRE, an issue which could be addressed by future theoretical research.

Indeed, the lack of evidence in support of models of belief learning in our experiment casts doubt on the validity of the view that players select myopic best responses to some beliefs concerning their opponents play, which are based on observations of their previous choices. This is indicated by the perverse coefficients on expected payoffs from different actions in many of the logit equations, where the expected payoffs are computed on the basis of alternative measures of empirical frequencies. In contrast, stimulus learning variables invariably have the right signs in all logit equations. Moreover, the fact that plays of the dominated strategy tended to be significant and bunched together towards the earlier stages of the game is not easily reconciled with explanations based on random preference shocks. The fact that dominated actions were used more often in games with higher payoff stakes, also runs against the view that the shocks represented ‘‘departures from rationality’’ owing to lapses of attention or to ‘‘mistakes.’’ A stimulus learning interpretation is more natural: players try different actions, and learn from their payoff realizations over time to not play the dominated action.

Nevertheless, it has been argued by Crawford (1995) that the evidence of the experiments with coordination games of van Huyck *et al.* (1990) can be explained by an adaptive belief learning process. Two comments are in order. First, there is a need to discriminate between belief learning and stimulus learning explanations of the nature of play in these games. Second, these coordination games characterized by multiple Pareto-ranked pure strategy equilibria may exhibit learning patterns that are quite different from constant sum games where players need to learn to play a unique mixed strategy equilibrium. As the various experiments of Cooper *et al.* (1990) with coordination games also exhibited, the nature of play typically converges quite rapidly to some pure strategy equilibrium—quite unlike the substantial divergences from equilibrium observed in constant sum games.[14] The games studied by Banks *et al.* (1994) or Schotter *et al.* (1994) in contrast involve extensive rather than normal forms; the former additionally introduce incomplete information. In these contexts also there appears to be a significant amount of non-equilibrium play, including plays of dominated strategies.[15] Banks *et al.* also report that plays of dominated strategy tend to die out in later periods, similar to our results. Evidence favoring an adaptive behavior hypothesis over a more sophisticated (refined) Nash equilibrium hypothesis was also obtained in a similar incomplete information setting by Brandts and Holt (1993).

A number of issues appear promising for future research. The ability of simple stimulus learning models to explain the pattern of play should be explored in a wide variety of experimental settings; in particular, their performance relative to belief learning models needs to be examined more widely. Theoretical research should also devote greater attention to the intermediate run and long run implications of such hypotheses. The wide variability of learning parameters suggests that we do not know much about the determinants of learning procedures employed. The extent to which the amount of learning differs with the level of experience needs to be understood better. Moreover, the possibility that players do not employ immutable learning rules, but modify the rules themselves with experience (as suggested by Selten, 1978), could be explored.

[14] Note that the coordination game experiments of Cooper *et al.* involved three choices for each player, while those of van Huyck *et al.* involved seven choices. On the other hand, substantial non-equilibrium play resulted in the experiments of Merlo and Schotter (1992) and Bull *et al.* (1987) on tournament games with a unique pure strategy equilibrium, where each player had 100 decisions to choose from. It is conceivable that the game dimension necessary to cause divergence from equilibrium predictions is higher in the case of games with pure rather than mixed strategy equilibria.

[15] In Schotter *et al.*, the play of the dominated strategy ranged from 12 to 20%, depending on the specific version. This is of a greater order of magnitude than in our experiment.

REFERENCES

Amemiya, T. (1981). "Qualitative Response Models: A Survey," *J. Econ. Lit.* **19**(4), 483–536.

Banks, J., Camerer, C., and Porter, D. (1994). "An Experimental Analysis of Nash Refinements in Signaling Games," *Games Econ. Behav.* **6**, 1–31.

Bendor, J., Mookherjee, D., and Ray, D. (1993). "Aspirations, Adaptive Learning and Cooperation in Repeated Games," working paper. Department of Economics, Boston University.

Boylan, R., and El-Gamal, M. (1993). "Fictitious Play: A Statistical Study of Multiple Economic Experiments," *Games Econ. Behavior* **5**, 205–222.

Brandts, J., and Holt, C. A. (1993). "Adjustment Patterns and Equilibrium Selection in an Experimental Signalling Game," *Int. J. Game Theory* **22**(3), 279–302.

Brown, J., and Rosenthal, R. (1990). "Testing the Minimax Hypothesis: A Re-examination of O'Neill's Game Experiment," *Econometrica* **58**, 1065–1082.

Bull, C., Schotter, A., and Weigelt, K. (1987). "Tournaments and Piece Rates; An Experimental Study," *J. Polit. Econ.* **95**, 1–33.

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

Cooper, R., DeJong, D., Forsythe, R., and Ross, T. (1990). "Selection Criteria in Coordination Games: Some Experimental Evidence," *Amer. Econ. Rev.* **80**(1), 218–233.

Cox J., Smith, V., and Walker, J. (1990). "Reward Saliency, Dominance and Decision Cost in Experimental Economics," mimeo. University of Arizona.

Crawford, V. (1995). "Adaptive Dynamics in Coordination Games," *Econometrica* **63**(1), 103–144.

Fudenberg D., and Kreps, D. (1993). "Learning Mixed Equilibria," *Games Econ. Behav.* **5**, 320–367.

Harsanyi, J. (1973). "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Equilibrium Points," *Int. J. Game Theory* **2**, 1–23.

Jordan, J. (1991). "Bayesian Learning in Normal Form Games," *Games Econ. Behav.* **3**, 60–81.

Jordan, J. (1993). "Three Problems in Learning Mixed Strategy Equilibria," *Games Econ. Behav.* **5**, 368–386.

Kreps, D., Milgrom, P., Roberts, J., and Wilson, R. (1982). "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma," *J. Econ. Theory* **27**, 245–252.

Maddala, G. S. (1986). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge Univ. Press.

McCabe K., Mukherji, A., and Runkle, D. (1993). "Learning and Limited Information in Games: An Experimental Study," mimeo. Federal Reserve Bank of Minneapolis.

McFadden, D. (1976). "Quantal Choice Analysis: A Survey," *Ann. Econ. Soc. Measure* **5**, 363–90.

McFadden, D. (1981). "Econometric Analysis of Probabilistic Choice," in *Structural Analysis of Discrete Data* (C. Manski and D. McFadden, Eds.). Cambridge, MA: MIT Press.

McKelvey, R., and Palfrey, T. (1992). "An Experimental Study of the Centipede Game," *Econometrica* **60**(4), 803–836.

McKelvey, R., and Palfrey, T. (1993). "Quantal Response Equilibria for Normal Form Games," mimeo. California Institute of Technology.

Merlo, A., and Schotter, A. (1992). "An Experimental Study of Learning in One and Two Person Games," Research Report 94-17. C. V. Starr Center, New York University.

Mookherjee, D. and Sopher, B. (1994). "Learning Behavior in an Experimental Matching Pennies Game," *Games Econ. Behav.* **7**, 62–91.

O'Neill, B. (1987). "Nonmetric Test of the Minimax Theory of Two-Person Zero Sum Games," *Proc. Nat. Acad. Sci. USA* **84**, 2106–2109.

Rapoport, A., and Boebel, R. (1992). "Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis," *Games Econ. Behav.* **4**, 261–283.

Roth, A., 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.

Schotter, A., Weigelt, K., and Wilson, C. (1994). "A Laboratory Investigation of Multiperson Rationality and Presentation Effects," *Games Econ. Behav.* **6**, 445–468.

Selten, R. (1978). "The Chain Store Paradox," *Theory Dec.* **9**, 127–159.

Selten, R. (1991). "Evolution, Learning and Economic Behavior," *Games Econ. Behav.* **3**, 3–24.

Selten, R., and Stoecker, R. (1986). "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames," *J. Econ. Behav. Organ.* **7**, 47–70.

Shapley, L. (1964). "Some Topics in Two-Person Games" in *Advances in Game Theory* (M. Dresher, L. Shapley, and A. Tucker, Eds.). Princeton, NJ: Princeton Univ. Press.

Smith, V., and Walker, J. (1993). "Monetary Rewards and Decision Cost in Experimental Economics," *Econ. Inquiry* **31**, 245–261.

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

van Huyck, J., Battalio, R., and Beil, R. (1990). "Tacit Coordination Games, Strategic Uncertainty and Coordination Failure," *Amer. Econ. Rev.* **89**(1), 234–248.