

Aspiration-Based and Reciprocity-Based Rules in Learning Dynamics for Symmetric Normal-Form Games

Dale O. Stahl

University of Texas at Austin

and

Ernan Haruvy

University of Texas at Dallas

Psychologically based rules are important in human behavior and have the potential of explaining equilibrium selection and separatrix crossings to a payoff dominant equilibrium in coordination games. We show how a rule learning theory can easily accommodate behavioral rules such as aspiration-based experimentation and reciprocity-based cooperation and how to test for the significance of additional rules. We confront this enhanced rule learning model with experimental data on games with multiple equilibria and separatrix-crossing behavior. Maximum likelihood results do not support aspiration-based experimentation or anticipated reciprocity as significant explanatory factors, but do support a small propensity for non-aspiration-based experimentation by random belief and non-reciprocity-based cooperation. © 2002

Elsevier Science (USA)

Key Words: aspiration; reciprocity; rules; learning; games.

In recent years, the field of learning in games has evolved considerably, with recent models able to make robust predictions in a variety of interesting games. Examples in the reinforcement learning arena include Erev & Roth (1998), Roth & Erev (1995), and Sarin & Vahid (1999). In belief learning, Fudenberg & Levine (1998) and Cheung & Friedman (1997, 1998) provide some comprehensive reviews. In hybrid models, Camerer & Ho (1998, 1999a, 1999b) have led the field. These models, though different in structure, are based on players adjusting, with noise and

Partial funding of this research was provided by Grants SBR-9410501 and SBR-9631389 from the National Science Foundation and the Malcolm Forsman Centennial Professorship, but nothing herein reflects their views.

Address correspondence and reprint requests to Dale O. Stahl, Department of Economics, University of Texas, Austin, TX 7812. E-mail: stahl@eco.utexas.edu.

inertia, in the direction of the best response to past observations. In that sense, players' sophistication is fairly limited. When hypotheses are permitted (in belief and hybrid models), players' hypotheses on others' frequencies of choice are limited to a weighted average of others' past frequencies. In reality, players are far more sophisticated and are able to reason iteratively, as Ho, Camerer & Weigelt (1998), Nagel (1995), and Stahl (1996) have shown. For example, a sophisticated player might believe that other players will tend to choose a best reply to the historical trend (which could be different from a weighted average of others' past frequencies) and hence choose a best reply to this belief. Action reinforcement dynamics cannot capture such behavior, since the behavior is not simply the tendency to choose an action. Rather the behavior is a function (or behavioral rule) that maps game information and histories into available actions.

Stahl's (1999, 2000a, 2000b, 2001) rule learning theory assumes that players probabilistically select behavioral rules and that the propensities to use specific rules are guided by the law of effect: the propensity of rules that perform better increase over time and vice versa. Precursors of this approach are Newell, Shaw, & Simon (1958). To operationalize the theory, a class of rules is assumed which includes the sophisticated rule mentioned above (see Section 2.3 for details). While this specification of the rule learning theory has been remarkably successful empirically (compared to other learning models), a shortcoming is that all the rules are either a form of trending or myopically rational in the sense of being (perhaps noisy) best replies to a current belief. We say this is a shortcoming because a growing number of laboratory experiments have documented that individuals deviate systematically from myopic self-interest behavior when there are possible future gains from influencing others. Specifically, such nonmyopic actions may be motivated by expectations of future reciprocity by others (e.g., Engle-Warnick & Slonim, 2001; Fehr & Gächter, 1998; Gneezy, Guth, & Verboven, 2001), attempts to teach others (e.g., Camerer, Ho, & Chong, 2000), or attempts to signal intended future actions in coordination games (e.g., Cooper, Dejong, Forsythe, & Ross, 1993).

For example, consider the coordination game presented in Fig. 1. The matrix of Fig. 1 gives the row player's payoffs, and the triangular graph depicts the aggregate choice frequencies of actions A and B for an experiment session. The dotted line depicts the separatrix between the area for which B is the best reply (upper area) and the area for which A is the best reply; C is a dominated action. The experiment session had 25 participants mean-matched for 12 periods, with population feedback after each period (see the Methods section for details). The first period is represented by the dot labeled with a "1", which lies in A's best-reply area. In the second and third periods (labeled "2" and "3" respectively), the choice frequencies move strongly in the direction of the best reply. Then in period 4 there is a surprising reversal of direction, followed in period 5 by the crossing to action B's best-reply area and eventual convergence to the payoff-dominant equilibrium (B). This separatrix-crossing phenomenon is robust (9 out of 20 sessions) and is incompatible with myopic rationality.

Thus, we have compelling reasons to develop better models that go beyond trending and myopically rational behavior to include other psychologically based modes of behavior. The primary purpose of this paper is to show how the rule

	A	B	C
A	20	0	60
B	0	60	0
C	10	25	25

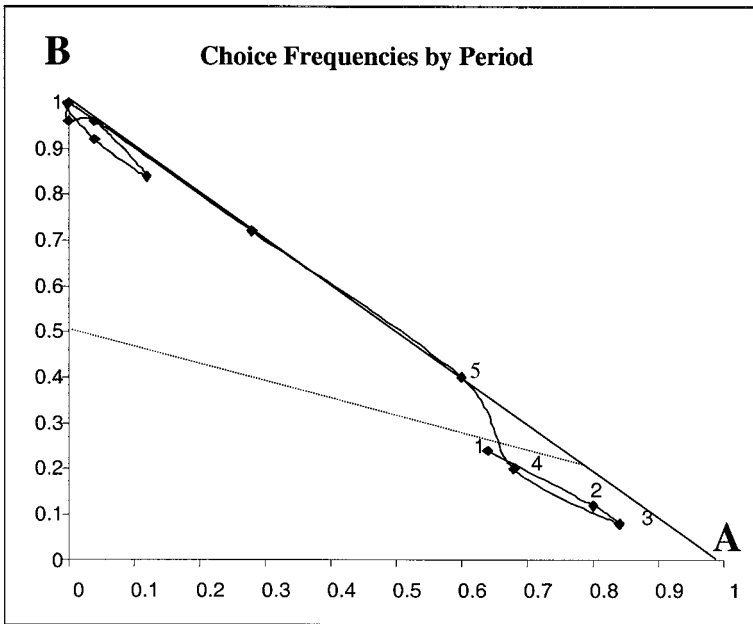


FIG. 1. Evidence of nonmyopically rational behavior.

learning theory can easily incorporate diverse psychologically based rules and how to test for the significance of additional rules. Having motivated the need with the example in Fig. 1, we will also use this example to motivate the behavioral rules that we will use to achieve our purpose. However, we will leave the mystery of the separatrix crossing unresolved, and we will not answer the quest for the best class of rules since that is not our primary goal. Nonetheless, the methodology we develop is a promising approach for future research.

Binmore & Samuelson (1997) offered a theoretical explanation for separatrix-crossing behavior based on aspirations and experimentation. Applied to the above example, players begin with aspirations of payoffs, which during the first three periods depicted in Fig. 1 are not fulfilled, inducing the players to experiment with other strategies. If experimentation leads to enough B (but not C) choices, then the separatrix can be crossed.

Another possible explanation of separatrix crossings is that some players make short-term sacrifices to teach others to coordinate on the payoff-dominant equilibrium (an interesting model of teaching behavior in repeated games is given in

Camerer *et al.*, 2000). We could think about these players as anticipating positive reciprocity from the other players. In general models of reciprocity (e.g., Dufwenberg & Kirchsteiger, 1998; Rabin, 1993), a costly action by one party which is favorable to another is due to intentional actions by the other party perceived to be favorable (positive reciprocity). Similarly, an unfavorable action is chosen in response to intentional actions perceived to be unfavorable (negative reciprocity). Though these responses may be costly for the individual, they could serve to enforce social norms, which may ultimately result in everyone being better off (Brandts & Charness, 1999; Fehr & Gächter, 1998; Hoffman, McCabe, & Smith, 1998).

We can incorporate Binmore & Samuelson's insight into the rule learning theory by introducing an *experimentation rule* whose propensity is sensitive to the difference between player aspirations and actual payoffs. Players start with an initial aspiration level and an initial propensity toward experimentation (loosely some random deviation from myopic best-reply behavior), and if the actual payoffs fall short of the aspiration level, the propensity to experiment increases, and vice versa. Allowing the aspiration level to be updated gradually could then explain separatrix-crossing behavior.

Alternatively, we can incorporate anticipated reciprocity into the rule learning theory by introducing a *cooperative rule* (in the current examples, to choose the symmetric payoff dominant action), an initial propensity to cooperate (i.e., use this rule), and an initial anticipated-reciprocity payoff (the payoff that would be obtained if everyone cooperates). The anticipated-reciprocity payoff would be updated only gradually, thereby representing patience in waiting for others to reciprocate. Positive reciprocal behavior by others would increase the propensity to cooperate, but if others failed to reciprocate, eventually the propensity to cooperate would plummet. We also consider an alternative modeling of anticipated reciprocity in the form of a tit-for-tat (TFT) rule.

The next section presents the basic rule learning model and the following section presents three enhancements. Following that, we present Methods, Results, and Discussion.

THE RULE LEARNING FRAMEWORK

The Game Environment

Consider a finite, symmetric, two-player game $G \equiv (N, A, U)$ in normal form, where $N \equiv \{1, 2\}$ is the set of players, $A \equiv \{1, \dots, J\}$ is the set of actions available to each player, U is the $J \times J$ matrix of expected utility payoffs for the row player, and U' , the transpose of U , is the payoff matrix for the column player. For notational convenience, let $p^0 \equiv (1/J, \dots, 1/J)'$ denote the uniform distribution over A .

We focus on single population situations in which each player is matched in every period with every other player; hence, the payoff relevant statistic for any given player is the probability distribution of the choices of the population, and this information is available to the players. To this end, p^t will denote the empirical frequency of all players' actions in period t . It is also convenient to define

$h^t \equiv \{p^0, \dots, p^{t-1}\}$ as the history of all players' choices up to period t with the novelty that p^0 is substituted for the null history. Thus, the information available to a representative player at the beginning of period t is $\Omega^t \equiv (G, h^t)$.

Behavioral Rules and the Rule Learning Dynamic

A *behavioral rule* is a mapping from information Ω^t to $\mathcal{A}(A)$, the set of probability measures on the actions A . For the purposes of presenting the abstract model, let $\rho \in R$ denote a generic behavioral rule in a space of behavioral rules R ; $\rho(\Omega^t)$ is the mixed strategy generated by rule ρ given information Ω^t .

The second element in the model is a probability measure over the rules: $\varphi(\rho, t)$ denotes the probability of using rule ρ in period t . Because of the nonnegativity restriction on probability measures, it is more convenient to specify the learning dynamics in terms of a transformation of φ that is unrestricted in sign. To this end, we define $w(\rho, t)$ as the *log-propensity* to use rule ρ in period t , such that

$$\varphi(\rho, t) \equiv \exp(w(\rho, t)) / \left[\int \exp(w(z, t)) dz \right]. \tag{1}$$

Given a space of behavioral rules R and probabilities φ , the induced probability distribution over actions for period t is

$$p(t) \equiv \int_R \rho(\Omega^t) d\varphi(\rho, t). \tag{2}$$

In other words, $p_j(t)$ is the probability that action j will be chosen in period t and is obtained by integrating the probability of j for each rule, $\rho_j(\Omega^t)$, with respect to the probability distribution over rules, $\varphi(\rho, t)$. Computing this integral is the major computational burden of this model.

The third element of the model is the equation of motion. The law of effect states that rules which perform well are more likely to be used in the future. This law is captured by the following dynamic on log-propensities,

$$w(\rho, t+1) = \beta_0 w(\rho, t) + g(\rho, \Omega^{t+1}), \quad \text{for } t \geq 1, \tag{3}$$

where $g(\cdot)$, is the reinforcement function for rule ρ conditional on information $\Omega^{t+1} = (G, h^{t+1})$, and $\beta_0 \in (0, 1]$ is an inertia parameter. We specify $g(\rho, \Omega^{t+1})$ as the rescaled expected utility that rule ρ would have generated in period t :

$$g(\rho, \Omega^{t+1}) = \beta_1 \rho(\Omega^t)' U p^t, \tag{4}$$

where $\beta_1 > 0$ is a scaling parameter.

While the ex ante rule propensities $\varphi(\rho, t)$ are the same for all individuals, diversity in the population is generated by independent draws from $\varphi(\rho, t)$ for each individual in each period. Since we are specifying a model of population averages rather than individual responses, the reinforcement function represents the cumulated effects of experience throughout the population.

Given a space of rules R and initial conditions $w(\rho, 1)$, the law of motion, Eq. (3), completely determines the behavior of the system for all $t > 1$. The remaining operational questions are (1) how to specify R , and (2) how to specify $w(\rho, 1)$.

An attractive feature of this general model is that it encompasses a wide variety of learning theories. For instance, to obtain replicator dynamics, we can simply let R be the set of J constant rules that always choose one unique action in A for all information states. Fictitious play and Cournot dynamics can be seen as very special cases in which R is a singleton rule which chooses a (possibly noisy) best-response to a belief that is a deterministic function of the history of play.¹ Moreover, the general model can include these constant rules, best-response rules, and other rules.

The Family of Evidence-Based Rules

Our approach to specifying the space of rules is to specify a finite number of empirically relevant discrete rules that can be combined to span a much larger space of rules.² In Stahl (1999, 2000a), the family of evidence-based rules was introduced as an extension of the Stahl & Wilson (1995) level- n rules. Evidence-based rules are derived from the notion that a player considers evidence for and against the available actions and tends to choose the action that has the most net favorable evidence based on the available information. Further, evidence is generated from three archetypal "theories of mind" (i.e., models of other players).

The first kind of evidence comes from a *null* model of the other players. The null model provides no reason for the other players to choose any particular action, so for the first period of play, by virtue of insufficient reason, the belief is that all actions are equally likely. The expected utility payoff to each available action given the null model is $y_{1j}(\Omega^1) \equiv Up^0$. We interpret y_{1j} as *evidence* in favor of action j stemming from the null model and no prior history.

For later periods ($t > 1$), the players have empirical data about the past choices of the other players. It is reasonable for a player to use simple distributed-lag forecasting: $(1 - \theta)p^0 + \theta p^1$ for period 2 with $\theta \in [0, 1]$. Letting $q^t(\theta)$ denote the forecast for period t and defining $q^0(\theta) \equiv p^0$, the following forecasting equation applies for all $t \geq 1$:

$$q^t(\theta) \equiv (1 - \theta)q^{t-1}(\theta) + \theta p^{t-1}. \quad (5)$$

The expected utility payoff given this belief is $y_{1j}(\Omega^t; \theta) \equiv Uq^t(\theta)$. We can interpret $y_{1j}(\Omega^t; \theta)$ as *level-1* evidence in favor of action j stemming from the null model and history h^t .

The second kind of evidence is based on the Stahl & Wilson (1995) *level-2* player who believes all other players are level-1 players and hence believes that the distribution of play will be $b(q^t(\theta))$, where $b(q^t(\theta)) \in \Delta(A)$ puts equal probability on all

¹ In fictitious play, the belief is the unweighted empirical frequency of observed choices, and in Cournot dynamics the belief is the most recent observed empirical frequency.

² Alternative approaches could use the production rules of ACT-R (Anderson, 1993), or fast and frugal heuristics (Gigerenzer, Todd, & the ABC Research Group, 1999).

best responses to $q'(\theta)$ and zero probability on all inferior responses. This is a second order theory of mind in which other minds are believed to be first order (i.e., level 1). Indeed, humans may have brain structures for this kind of representation (Churchland & Sejnowski, 1992). The expected utility conditional on this belief is $y_2(\Omega'; \theta) \equiv Ub(q'(\theta))$. We can interpret $y_{2j}(\Omega'; \theta)$ as level-2 evidence in favor of action j .

The third theory of mind is Nash equilibrium theory. Letting p^{NE} denote a Nash equilibrium of G , $y_3 = Up^{NE}$ provides a third kind of evidence on the available actions. For games with multiple NE, the evidence for each NE is equally weighted.³

So far we have defined three kinds of evidence: $Y \equiv \{y_1, \dots, y_3\}$. The next step is to weigh this evidence and specify a probabilistic choice function. Let $v_k \geq 0$ denote a scalar weight associated with evidence y_k . We define the weighted evidence vector

$$\bar{y}(\Omega'; v, \theta) \equiv Y(\Omega'; \theta) v, \tag{6}$$

where $v \equiv (v_1, \dots, v_3)'$.

There are many ways to go from such a weighted evidence measure to a probabilistic choice function. We opt for the multinomial logit specification because of its computational advantages when it comes to empirical estimation. The implicit assumption is that the player assesses the weighted evidence with some error and chooses the action that from his or her perspective has the greatest net favorable evidence. Hence, the probability of choosing action j is

$$\hat{p}_j(\Omega'; v, \theta) \equiv \exp[\bar{y}_j(\Omega'; v, \theta)] / \sum_{\ell} \exp[\bar{y}_{\ell}(\Omega'; v, \theta)]. \tag{7}$$

Note that, given the four-dimensional parameter vector (v, θ) , Eq. (7) defines a mapping from Ω' to $\Delta(A)$, and hence $\hat{p}(\bullet; v, \theta)$ is a behavioral rule as defined abstractly above. By putting zero weight on all but one kind of evidence, Eq. (7) defines an archetypal rule—one for each kind of evidence corresponding to the underlying model of other players. The four-dimensional parameter vector (v, θ) generates the space of rules spanned by these archetypal rules. To show the behavioral effect of various v values, Table 1 gives the predicted initial-period choice frequencies for games 13 and 16 for several values of (v_1, v_2, v_3) .

Next, we represent behavior that is random in the first period and follows the herd in subsequent periods. Following the herd does not mean exactly replicating the most recent past, but rather following the past with inertia as represented by $q'(\theta)$. Hence, Eq. (5) represents herd behavior as well as the beliefs of level-1 types.

Finally, we allow for uniform trembles by introducing the uniformly random rule. Thus, the base model consists of a four-dimensional space of evidence-based rules (v, θ) , a herd rule characterized by θ , and uniform trembles.

³ See Haruvy & Stahl (1999).

TABLE 1

Choice Frequencies for Various Evidence-Based Rules

(v_1, v_2, v_3)	A	B	C
Game 13			
(0.23, 0, 0)	0.760	0.076	0.164
(1.08, 0, 0)	1.0	0	0
(0.81, 0.27, 0)	0.937	0	0.063
(0.81, 0, 0.23)	1.0	0	0
Game 16			
(0.23, 0, 0)	0.659	0.151	0.151
(1.08, 0, 0)	1.0	0	0
(0.81, 0.27, 0)	1.0	0	0
(0.81, 0, 0.23)	0.492	0.492	0.017

The Initial Distribution of Log-Propensities

Letting δ_h denote the initial probability of the herd rule, and letting ε denote the initial probability of the trembles, we set $w(\rho_{\text{herd}}, 1) = \ln(\delta_h)$ and $w(\rho_{\text{tremble}}, 1) = \ln(\varepsilon)$. The initial log-propensity function for the evidence-based rules is specified as

$$w(v, \theta, 1) \equiv -0.5 \|(v, \theta) - (\bar{v}, \bar{\theta})\|^2 / \sigma^2 + A, \quad (8)$$

where $\|(v, \theta) - (\bar{v}, \bar{\theta})\|$ denotes the distance⁴ between rule (v, θ) and the mean of the distribution $(\bar{v}, \bar{\theta})$, and A is determined by the requirement that Eq. (1) integrated over the space of evidence-based rules is exactly $1 - \delta_h - \varepsilon$. Hence, the initial propensity over the evidence-based rules is essentially a normal distribution with mean $(\bar{v}, \bar{\theta})$ and standard deviation σ .

Transference

Since this theory is about rules that use game information as input, we should be able to predict behavior in a temporal sequence that involves a variety of games. For instance, suppose an experiment consists of one run with one game for T periods, followed by a second run with another game for T periods. How is what is learned about the rules during the first run transferred to the second run with the new game? A natural assumption would be that the log-propensities at the end of the first game are simply carried forward to the new game. Another extreme assumption would be that the new game is perceived as a totally different situation so the log-propensities revert to their initial state. We opt for a convex combination

$$w(\rho, T+1) = (1-\tau) w(\rho, 1) + \tau w(\rho, T^+), \quad (9)$$

⁴ This is a log-linear Euclidian distance: the v variable is measured on a logarithmic scale, while θ is on a linear scale; see Appendix A1 of Stahl (2001) for computational details.

where T^+ indicates the update after period T of the first run, and τ is the *transference* parameter. If $\tau = 0$, there is no transference, so period $T + 1$ has the same initial log-propensity as period 1; and if $\tau = 1$, there is complete transference, so the first period of the second run has the log-propensity that would prevail if it were period $T + 1$ of the first run (with no change of game). This specification extends the model to any number of runs with different games without requiring additional parameters.

The Likelihood Function

The theoretical model involves 10 parameters: $\beta \equiv (\bar{v}_1, \bar{v}_2, \bar{v}_3, \bar{\theta}, \sigma, \delta_h, \varepsilon, \beta_0, \beta_1, \tau)$. The first four parameters $(\bar{v}_1, \bar{v}_2, \bar{v}_3, \bar{\theta})$ represent the mean of the population's initial propensity $w(\rho, 1)$ over the evidence-based rules, and σ is the standard deviation of that propensity; the next two parameters (δ_h, ε) are the initial propensities of the herd and tremble rules, respectively; β_0 and β_1 are the learning parameters of Eqs. (3) and (4); and τ is the transference parameter in Eq. (9) for the initial propensity of the subsequent runs. Substituting Eq. (7) into Eq. (2), the rule propensities and law of motion yield population choice probabilities:

$$p_j(t | \beta) = \int_R \hat{p}_j(\Omega^t; v, \theta) \varphi(v, \theta, t | \beta) d(v, \theta). \quad (10)$$

To compute this integral, a grid was placed on the (v, θ) -space of rules (see Appendix A1 of Stahl, 2001 for details). Let n_j^t denote the number of participants who choose action j in period t . Then the logarithm of the joint probability of the data conditional on β is

$$LL(\beta) \equiv \sum_t \sum_j n_j^t \log[p_j(t | \beta)]. \quad (11)$$

To find a β vector that maximizes $LL(\beta)$, we use a simulated annealing algorithm (Goffe, Ferrier, & Rogers, 1994) for high but declining temperatures and then feed the result into the Nelder & Mead (1965) algorithm. The simulated annealing algorithm is effective in exploring the parameter space and finding neighborhoods of global maxima but very slow to converge, while the latter algorithm converges much faster.

ENHANCEMENTS TO THE RULE LEARNING MODEL

Aspiration-Based Experimentation

Two components of an aspiration-based experimentation rule are the experimentation act and the aspiration level. We let $p^x(t)$ denote the probability distribution of the experimentation act in period t . Two obvious candidates for $p^x(t)$ are the uniform distribution p^0 for experimentation by *random choice*, and the recent past p^{t-1} for experiment by *imitation*. The third alternative we will consider is experimentation by *random beliefs* in which $p^x(t)$ is the proportion of the probability

simplex for which each choice is the best response. Such choice probabilities would arise from beliefs that are uniformly distributed over the probability simplex.

Let $a^x(t)$ denote the aspiration level for period t . We assume an adaptive expectations dynamic for the aspiration level

$$a^x(t+1) = \theta_2 a^x(t) + (1 - \theta_2) p^x(t)' U p^t \quad (12)$$

with initial condition $a^x(1)$. With $0 < \theta_2 < 1$, the aspiration level will adjust gradually to the actual expected utility that the experimentation act would have generated each period. If initial play is diverse and the experimentation act is diverse, then $p^x(t)' U p^t$ will be highly correlated with the population average payoff. Hence, if the latter is much less than the initial aspiration level and if $\theta_2 < 1$, then the aspiration level will gradually decline but remain above the population average payoff, and this positive difference will increase the likelihood that the experimentation rule will be used. Conversely, if the population average payoff is higher than the initial aspiration level, then the propensity to experiment will decline (don't rock the boat).

The rule learning framework can represent this dependence of the propensity to experiment on the aspiration level by using $\beta_1 a^x(t+1)$ as the reinforcement, $g(\rho_x, \Omega^{t+1})$, for the experimentation rule in Eq. (3). If initial aspirations are disappointed, then the log-propensity to experiment will increase, and vice versa. Asymptotically, as long as $\theta_2 < 1$, Eq. (12) has the property that the long-run propensity to use the experimentation act, $p^x(t)$, will depend on the actual expected payoff from that act, since initial aspirations are geometrically discounted. Conversely, the larger θ_2 , the longer will be the effect of the initial aspiration level. For the game in Fig. 1, simulations of rule learning with an initial aspiration level of 50 or higher and a value of $\theta_2 > 0.9$ produce separatrix crossings.

In addition to specification of $p^x(t)$, we need to specify an initial aspiration level, $a^x(1)$, and an initial propensity, δ_x , to experiment (i.e., $\varphi(\rho_x, 1) = \delta_x$). It is natural to set $a^x(1)$ equal to the mean of the range of payoffs in the game: $[\max_{jk} U_{jk} - \min_{jk} U_{jk}]/2$. We let δ_x be a free parameter to be estimated from the data. We further assume that the transference equation applies to the aspiration level, so

$$a^x(T+1) = (1 - \tau) a^x(1) + \tau a^x(T^+). \quad (9')$$

While this model of aspiration-based experimentation will suffice to illustrate the richness of the rule learning theory, it does not exhaust all possible models of aspiration-based experimentation—which would go beyond the purpose of this paper.

Reciprocity-Based Cooperation

For the class of symmetric normal-form games, we define the *symmetric cooperative action* as the action with the largest diagonal payoff, and we restrict attention to the subclass of games with a unique symmetric cooperative action. Let p^c denote the degenerate probability distribution that puts unit mass on this cooperative

action; we will refer to this as the *cooperative rule*, ρ_c . But why would a rational player choose the cooperative rule?

A rational player who believes that almost everyone else will select the cooperative action would anticipate a payoff equal to the maximum diagonal payoff if he or she chooses the cooperative action. Then, if the cooperative action is a best response to this belief (i.e., if it is a payoff-dominant Nash equilibrium), the rational player will choose the cooperative action. If the cooperative action is not a best-reply to itself, a myopically rational player will not choose it, but a forward-looking rational player could still choose it if he or she believes that others will reciprocate leading to a higher payoff than the path that would follow from myopic best replies. If others do not cooperate at the start, the realized payoff to cooperation will be much less than anticipated, and so the player's belief about the likelihood of reciprocation may decline (especially if the player is impatient).

Let θ_2 denote the level of patience of such a player, so the anticipated payoff from cooperation, $a^c(t)$, (via future reciprocity) follows the simple adaptive expectations dynamic

$$a^c(t+1) = \theta_2 a^c(t) + (1 - \theta_2) p^c U p^c, \quad (13)$$

where $a^c(1) = p^c U p^c$, the largest diagonal payoff. This initial anticipated payoff is geometrically discounted over time; so provided $\theta_2 < 1$, the anticipated payoff from cooperation will tend toward a geometrically weighted average of actual expected payoffs to cooperation. The larger θ_2 , the slower the adjustment, so if the propensity to cooperate is positively correlated with $a^c(t)$, cooperative efforts could persist despite the lack of reciprocation.

The rule learning framework can represent this dependence of the propensity to cooperate on anticipated reciprocity by using $\beta_1 a^c(t+1)$ as the reinforcement, $g(\rho_c, \Omega^{t+1})$, for the cooperative rule in Eq. (3). When $\theta_2 = 1$, we would have a stubborn, irrational cooperative rule, but otherwise, the propensity to cooperate will depend on the difference between the anticipated payoff and the actual expected payoffs of alternative rules. When cooperation is reciprocated, so its payoff stays high relative to alternative rules (which recommend different actions), then the propensity to cooperate will increase, and vice versa.

We further assume that the transference equation applies to the anticipated payoff, so

$$a^c(T+1) = (1 - \tau) a^c(1) + \tau a^c(T^+). \quad (9'')$$

Finally, we let δ_c denote the initial propensity to cooperate (i.e., $\varphi(\rho_c, 1) = \delta_c$).

While this model of reciprocity-based cooperation will suffice to illustrate the richness of the rule learning theory, it does not exhaust all possible models of reciprocity-based cooperation—which would go beyond the purpose of this paper. Nonetheless, we consider one more model in the next section.

A Tit-for-Tat Rule

TFT behavior gained prominence following Axelrod's (1981, 1984) experimental prisoner's dilemma tournaments in which TFT emerged as the winning strategy,

and Kreps, Milgrom, Roberts, & Wilson's (1982) rational underpinning of TFT behavior in the finitely repeated prisoner's dilemma with incomplete information. In a two-player prisoner's dilemma, the TFT rule specifies cooperation in the first period and then mimicking the most recent action of one's opponent thereafter. The TFT rule we examine here begins cooperatively and thereafter cooperates if a threshold number of players have cooperated in the most recent period; otherwise it chooses a best reply to the recent period. Formally, let $f_c(t)$ denote the fraction of the population that cooperated in period t , and let $c \in (0, 1)$ denote the threshold. Then, we define the TFT probability of cooperating in period $t + 1$ as

$$\psi(t+1; f_c(t), c) \equiv \begin{cases} 0.5[f_c(t)/c]^\gamma, & \text{if } f_c(t) \leq c, \\ 1 - 0.5\{[1 - f_c(t)]/[1 - c]\}^\gamma, & \text{if } f_c(t) > c, \end{cases} \quad (14)$$

where $\gamma \geq 1$ is a precision parameter to be estimated along with c . This probability has the property that (i) it is 0 when $f_c(t) = 0$, (ii) it is 1 when $f_c(t) = 1$, (iii) it is 0.5 when $f_c(t) = c$, and (iv) it is S-shaped and monotonically increasing in $f_c(t)$. For large values of γ , the probability function is close to a step function at $f_c(t) = c$. When $c \approx 0$, the cooperative action is chosen for almost all realizations of $f_c(t)$, while when $c \approx 1$, the noncooperative best-reply action is chosen for almost all realizations of $f_c(t)$.

From Eq. (14), it is straightforward to define the TFT behavioral rule $\rho_{\text{TFT}}(\Omega^{t+1})$ which plays the cooperative action with probability $\psi(t+1; f_c(t), c)$ and plays the best reply to p^t with probability $1 - \psi(t+1; f_c(t), c)$. In accordance with Eq. (4), the reinforcement function $g(\rho_{\text{TFT}}, \Omega^{t+1}) = \beta_1 \rho_{\text{TFT}}(\Omega^t)' U p^t$. To complete the rule learning model with this TFT rule, we specify the initial propensity to use the TFT rule as δ_{TFT} .

METHODS

Participants

We conducted 25 sessions, each involving 25 different participants. All participants were inexperienced in this or similar experiments, and all were upper division undergraduates or noneconomics graduate students. Two sessions were conducted at Texas A&M University, and the rest were conducted at the University of Texas.

Apparatus

Participants were seated at private computer terminals separated so that no participant could observe the choices of other participants. The relevant game, or decision matrix, was presented on the computer screen. Each participant could make a choice by clicking the mouse button on any row of the matrix, which then became highlighted. In addition, each participant could make hypotheses about the choices of the other players. An on-screen calculator would then calculate and display the hypothetical payoffs to each available action given each hypothesis. Participants were allowed to make as many hypothetical calculations and choice

revisions as time permitted. Following each time period, each participant was shown the aggregate choices of all other participants and could view a record screen with the history of the aggregate choices of other participants for the entire run.

Design and Procedures

Each experiment session consisted of two runs of 12 periods each. In the first run, a single 3×3 symmetric game was played for 12 periods, and in the second run, a different 3×3 symmetric game was played for 12 periods.

A mean-matching protocol was used. In each period, a participant's token payoff was determined by his or her choice and the percentage distribution of the choices of all other participants, $p(t)$, as follows: The row of the payoff matrix corresponding to the participant's choice was multiplied by the vector of choice distribution of the other participants. Token payoffs were in probability units for a fixed prize of \$2.00 per period. In other words, the token payoff for each period gave the percentage chance of winning \$2 for that period. The lotteries that determined final monetary payoffs were conducted following the completion of both runs using dice. Specifically, a random number uniformly distributed on $[00.0, 99.9]$ was generated by the throw of three 10-sided dice. A player won \$2.00 if and only if his or her token payoff exceeded his or her generated dice number. Payment was made in cash immediately following each session.

The Games

We select 10 games (see Fig. 2) from those investigated in Haruvy & Stahl (1999), with properties that make them ideal for a thorough study of learning dynamics. These are games 1, 4, 9, 12, 13, 14, 16, 18, and 19 of Haruvy & Stahl (1999). In addition, we have a 3×3 version of a 5×5 game used in Stahl (1999), for which cooperative behavior seemed to persist (labeled game 21).

What makes these games ideal? We begin by noting that games 1, 4, 13, 14, 16, and 19 are characterized by multiple equilibria which are Pareto ranked. The selection principles of payoff dominance, risk dominance, and security do not all coincide in any of these six games, and no pair of principles coincides in all of these six games. Most important the payoff dominant equilibrium action never coincides with the level-1 action of Stahl & Wilson (1994, 1995). This is important given the prediction by Haruvy & Stahl (1999) that level-1 is a heavy component in the determination of initial conditions. According to the predictions generated by the Haruvy & Stahl (1999) model, initial conditions in these six games will generally fall in a basin of attraction (for popular models of adaptive dynamics) inconsistent with the payoff-dominant choice.

Games 1, 14, and 19 begin with initial conditions far from uniform and very little movement is observed thereafter, with the exception of one run of game 14, where a significant proportion of the group goes against their best response to cross the separatrix toward the payoff dominant Nash equilibrium. Game 13 has a long dynamic path from initial conditions to the final outcome. The final period outcomes of Game 16 are almost equally split between the two equilibria, although the initial conditions fall usually in the non-payoff-dominant basin, hence resulting in

1	A	B	C
A	70	60	90
B	60	80	50
C	40	20	100

4	A	B	C
A	70	30	20
B	60	60	30
C	45	45	40

9	A	B	C
A	30	50	100
B	40	45	10
C	35	60	0

12	A	B	C
A	30	100	22
B	35	0	45
C	51	50	20

13	A	B	C
A	60	60	30
B	30	70	20
C	70	25	35

14	A	B	C
A	50	0	0
B	70	35	35
C	0	25	55

16	A	B	C
A	20	0	60
B	0	60	0
C	10	25	25

18	A	B	C
A	25	30	100
B	60	31	51
C	95	30	0

19	A	B	C
A	80	60	50
B	60	70	90
C	0	0	100

21	A	B	C
A	68	4	49
B	86	41	4
C	72	25	39

FIG. 2. Games.

separatrix crossings. Game 18 was investigated because its theoretical dynamics involve a persistent cycle. Games 9 and 12 have only a mixed-strategy equilibrium and hence potentially interesting dynamic paths.

We also investigate three additional versions of game 13 (with 20 added to each cell, "13+"; with 20 subtracted from each cell, "13-"; and with payoffs rescaled to $[0, 100]$; "13r"), and an additional version of game 16 (with 20 added to each cell, "16+"). The motivation for this variation was that if the initial aspiration level, $a^x(1)$, is fixed independent of the game payoffs (for example, determined by the participant's anticipated compensation for the whole experimental session), then adding 20 to all payoffs should change behavior. For instance, if $a^x(1) = 50$, then in game 16 the "A" payoff of 20 is substantially less than the initial aspiration level

TABLE 2

List of Games by Experimental Session

Date	First Run	Second Run
2/17/98	16	9
2/19/98	19	12
4/7/98	14	1
4/9/98	13	21
6/18/98	16	13
7/16/98	16+	13-
7/30/98	16+	13r
1/28/99	16	19
3/2/99	16+	19
6/9/99	16	13
6/10/99	16+	13
8/6/99	16	13+
8/11/99	16+	13+
1/27/00	16	14
2/16/00	16+	14
2/24/00	16	1
4/13/00	16+	14
7/13/00	16	4
7/19/00	16	14
7/20/00	16+	1
7/26/00	16	19
7/27/00	16+	1
8/2/00a	16+	18
8/2/00b	19	4
8/3/00	1	4

thereby inducing experimentation, while with 20 added to each cell, since then 40 is much closer to 50, experimentation (and hence separatrix crossings) should occur less often. The experimental evidence, however, is 5 and 4 crossings out of ten sessions of each version of game 16 respectively—a statistically insignificant difference at all commonly accepted significance levels. We therefore discard the hypothesis of a fixed initial aspiration level independent of the game payoffs.

The games used in each session are listed in Table 2. The data from the first four sessions were analyzed in Stahl (2000a, 2000b, 2001). The data from the other 21 sessions have not been analyzed before. The entire data set is available upon request from the authors.

Note that the range of payoffs for the games in Fig. 2 differs substantially from $[20, 70]$ to $[0, 100]$. In Haruvy & Stahl (1999), we found support for the hypothesis that a single payoff rescaling parameter, β_1 , applies to a wide range of games provided the payoffs are first rescaled to $[0, 100]$. On the current data set, we also found that the maximized LL increased with such rescaling. Therefore, in this paper we report only results obtained from rescaling to $[0, 100]$. This implies that all the

TABLE 3
Individual Choices for Session 6/18/1998, Run 1 (Game 16)

		SUBJECT																								
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
PERIOD	1	C	A	A	A	B	A	A	B	A	B	B	A	B	C	A	A	A	C	A	A	A	B	A	A	A
	2	A	A	A	A	A	A	A	C	A	A	A	A	C	A	A	B	A	A	A	A	B	A	B	A	A
	3	B	A	A	A	A	A	B	C	A	A	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A
	4	B	A	A	A	A	C	B	C	A	A	B	A	A	C	B	A	A	A	A	A	A	A	A	B	A
	5	A	B	A	A	A	B	B	A	A	A	B	A	B	B	A	B	A	A	A	A	B	A	A	B	B
	6	A	B	B	A	A	B	B	B	A	B	B	B	B	B	A	B	B	B	B	A	B	B	B	B	B
	7	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
	8	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	B	B	B	B	B	B	B	C	B
	9	B	B	C	B	B	B	B	B	A	B	B	B	B	B	A	B	B	B	B	B	B	B	B	A	B
	10	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A
	11	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	C	B
	12	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B

versions of games 13 and 16 are “equivalent,” an assumption that has been verified by independent analysis of the choice behavior across the different versions.⁵

RESULTS

In addition to the plot of the dynamic path for one of the sessions shown in Fig. 1, we also report the individual choice behavior for that data in Table 3. A cursory look at Table 3 suggests that participants, for the most part, do not persist in cooperative behavior (action B in that game) unless others are also simultaneously cooperative. For instance, although six participants chose B in period 1 (putting the initial play close to the separatrix), all but one abandoned cooperation in period 2, and none of the three who cooperated in period 2 continued to do so in period 3. Even in the critical period 4, of the five who chose B, only two persisted with B in period 5. While nine of the ten participants who cooperated in period 5 continued to cooperate in period 6, by then B was the myopic best response. Similar patterns can be observed throughout the data. Hence, an ocular analysis of the data favors the experimentation explanation over the anticipated reciprocity explanation. We will rigorously test this and other hypotheses next.

The Basic Rule Learning Model.

We estimated the 10 parameters of the basic rule learning model using all the data from the 25 sessions. We pool over games in order to try to identify regular features of learning dynamics that are general and not game-specific, so we can be more confident that these features will be important in predicting out-of-sample

⁵This rescaling implies that the initial aspiration level in rescaled units is $a^*(1) = [\max_{jk} U_{jk} - \max_{jk} U_{jk}]/2 = 50$, for all games, which implies that the initial aspiration level in un-rescaled units varies across games depending on the original game payoffs.

TABLE 4

Parameter Estimates of Basic Rule Learning Model

	Stahl (2000b)	Current data set
\bar{v}_1	0.797 (0.207)	0.376 (0.201)
\bar{v}_2	0.0696 (0.021)	0.000 (10^{-8})
\bar{v}_3	0.000 (10^{-8})	0.000 (10^{-7})
$\bar{\theta}$	0.647 (0.026)	0.938 (0.049)
σ	0.767 (0.086)	0.800 (0.104)
δ_h	0.533 (0.028)	0.441 (0.021)
ε	0.090 (0.023)	0.123 (0.015)
β_0	1.000 (0.002)	1.000 (0.003)
β_1	0.00790 (0.00145)	0.00465 (0.00097)
τ	1.000 (0.049)	0.308 (0.145)
LL		-8550.21

behavior. The ML parameter estimates and standard errors⁶ are given in Table 4, and the maximized LL is -8550.21 .

There are similarities and differences between these parameter estimates and those for a different data set reported in Stahl (2000b).⁷ First, as in Stahl (2000b), we find $\beta_0 = 1$, so past propensity differences between rules persist unless affected by performance. The initial propensities for herd behavior (δ_h) and trembles (ε), and hence the total propensity to use some evidence-based rule, are similar. Also the standard deviation of the initial distribution over evidence-based rules, σ , is similar, but the mean (\bar{v} , $\bar{\theta}$) is notably different. \bar{v}_2 is zero and \bar{v}_1 is less. Despite the low t -ratio for \bar{v}_1 , the lower bound on the 95% confidence interval was 0.296, which is still large enough to put substantial probability on the best response to the level-1 belief. However, \bar{v}_3 is still precisely zero. $\bar{\theta}$ is much higher, implying that beliefs adjust quicker, while the rescaling parameter for the reinforcement function shrinks 38%. However, the most substantial difference is the transference parameter that was estimated to be 1 in Stahl (2000b), but only 0.308 here; also the standard error is much larger.

Despite these differences, the contribution of rule learning is robust. Restricting $\beta_0 = 1$ and $\beta_1 = 0$ yields a model with diverse rules but no rule learning. The maximized LL decreases to -8569.93 . Since the transference parameter (τ) no longer plays a role, the likelihood-ratio test has three degrees of freedom, and the hypothesis of no-rule-learning has a p -value of less than 3×10^{-9} .

To illustrate the amount of rule learning that takes place (hence the behavioral effect of β_1), Fig. 3 shows selected $\varphi(v, \theta, t | \beta)$ values aggregated over the domain of θ for the initial period ($t = 1$), after the 12 periods of the first run ($t = 12$), and after the 12 periods of the second run ($t = 24$) for a typical session (namely 6/18/98

⁶ Standard errors and confidence intervals were obtained via parametric bootstrapping (Efron & Tibshirani, 1993, pp. 53–55, 169–177).

⁷ To ensure stability of the dynamic process, an upper bound of 1 is imposed on β_0 .

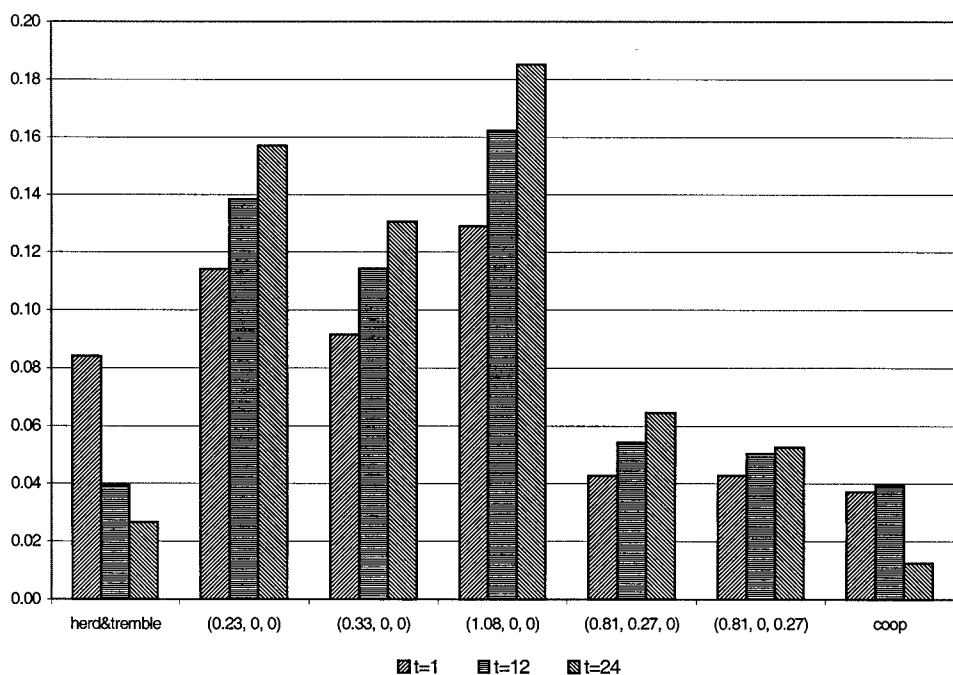


FIG. 3. $\varphi(t)$ for 6/18/98 session.

for which the first-run data are given in Fig. 1 and Table 3). These φ values come from the model of the Reciprocity-Based Cooperation Results section. Seven rule categories are shown, accounting for over 96% of the probability mass. The “herd&tremble” category on the left aggregates the herd rules and the tremble rule and shows a substantial decline in the propensities to use these rules. The next five rules are evidence-based rules where the three numbers that label each category are the values of (v_1, v_2, v_3) ;⁸ thus, the first three categories are level-1 rules with increasing degrees of precision. These rules show a substantial increase in propensity over time. The next rule combines level-1 and level-2 evidence, and the next rule combines level-1 and Nash evidence; both rules show a modest increase in propensity over time. The final category is a pure cooperation rule, the propensity of which appears fairly constant over the first run and then declines dramatically over the second run.

Aspiration-Based Experimentation Results

We investigated three versions of experimentation rules: (1) uniformly random choice, (2) imitation of the recent population distribution, and (3) uniformly random beliefs. The parameter estimates, standard errors, and maximized LL for each version are given in Table 5. In terms of LL, experimentation by imitation has the lowest LL (-8549.69 ; a statistically insignificant improvement of only 0.52).

⁸ Note that these are the same values as used in Table 1.

TABLE 5
Parameter Estimates of Enhanced Models

	Aspiration-based experimentation			Anticipated reciprocity
	Random choice	Imitation	Random beliefs	
\bar{v}_1	0.361 (0.195)	0.393 (0.241)	0.324 (0.061)	0.324 (0.160)
\bar{v}_2	0.000 (10^{-7})	0.000 (0.002)	0.0067 (0.007)	0.000 (0.0004)
\bar{v}_3	0.000 (10^{-3})	0.000 (0.002)	0.000 (0.0005)	0.000 (0.0003)
$\bar{\theta}$	0.920 (0.051)	0.908 (0.057)	0.923 (0.044)	0.859 (0.043)
σ	0.747 (0.108)	0.828 (0.105)	0.587 (0.111)	0.715 (0.115)
δ_h	0.448 (0.022)	0.258 (0.143)	0.426 (0.021)	0.430 (0.023)
δ_x	0.108 (0.016)	0.171 (0.133)	0.0329 (0.014)	0.037 (0.006)
ε	0.000 (0.002)	0.131 (0.018)	0.115 (0.013)	0.084 (0.014)
θ_2	0.757 (0.186)	0.588 (0.352)	0.860 (0.321)	0.000 (0.001)
β_0	1.000 (0.003)	1.000 (0.003)	1.000 (0.003)	1.000 (0.002)
β_1	0.00500 (0.00102)	0.00486 (0.00099)	0.00558 (0.00100)	0.00328 (0.00067)
τ	0.208 (0.168)	0.290 (0.132)	0.213 (0.136)	0.888 (0.179)
LL	-8546.83	-8549.69	-8544.70	-8522.41

This poor performance is due to the fact that the basic rule learning model already includes imitation via the herd rules. We also find that the estimates for $(\delta_h, \delta_x, \varepsilon)$ for the imitation version are quite different from any other version, which is attributable to multicollinearity between the imitation rule and the herd rules. Note the low t -ratio for δ_x , indicating that we cannot reject the hypothesis that $\delta_x = 0$.

Experimentation by uniformly random choice shows a LL gain of 3.38; with two degrees of freedom this gain has a p -value of 0.034, which is not impressive considering there are 625 participants in the data set. Moreover, the (δ_x, ε) estimates are quite different from the basic rule learning model, which is probably attributable to the multicollinearity between random-choice experimentation and the tremble rule in the basic rule-learning model. Despite the seemingly respectable t -ratio for θ_2 , the 95% confidence interval is (0.0007, 0.921), shedding doubt on the role of aspirations in this model.

The largest gain in LL is for experimentation by random beliefs. This gain of 5.51 has a p -value of 0.004 and so is statistically significant. Again, we find a large confidence interval for θ_2 (0.0001, 1.0), shedding doubt on the role of aspirations. We can test the hypothesis that aspirations are unimportant by setting $\theta_2 = 0$; then the initial aspiration level is fully discounted so the experimentation act, $p^x(t)$, is reinforced by its actual expected payoff to the most recent population choices—like all other behavioral rules in the model. The maximized LL of this restricted model decreases by only 1.10 and hence is not statistically significant at any common acceptance level. Therefore, we cannot reject the hypothesis that experimentation by random beliefs does not depend on aspirations in this data set. Yet, there is support for a simple experimentation-by-random-belief rule (3.3%) that is reinforced like any other behavioral rule. Since none of the other versions of aspiration-based experimentation does substantially better, we conclude more generally that aspirations (as modeled) play essentially no role in this data set.

Reciprocity-Based Cooperation Results

Table 5 also gives the parameter estimates, standard errors, and maximized LL for the anticipated reciprocity model. We note that the estimate of the patience parameter, θ_2 , is zero, indicating that players immediately react to the payoff from the cooperative action rather than reinforcing cooperative propensities by anticipated future reciprocity. The individual choice behavior (e.g., Table 3) is consistent with this finding. On the other hand, the estimated initial propensity to cooperate (δ_c) is about 3.7% and statistically different from 0 (p -value $< 10^{-13}$). Therefore, we can strongly reject the hypothesis that there is no cooperative behavior in this data set (i.e., $\delta_c = 0$). Another curious finding is that the transference estimate increases to 0.888.⁹

The gain in LL due to a non-reciprocity-based cooperative rule is 27.81, which is much larger than the gain from rule learning without any cooperative rule (19.72). Out of curiosity, we estimated the model with a non-reciprocity-based cooperative rule but without rule learning ($\beta_1 = 0$) and obtained a LL value of -8550.69 ; hence, adding a non-reciprocity-based cooperative rule to the no-rule-learning models yields a LL gain of 19.24, while the further addition of rule learning yields a LL gain of 28.28. Therefore, we can even more strongly reject the hypothesis of no rule learning in the presence of this cooperative rule.

Because it is irrational to anticipate reciprocity in the last period of an experiment, we considered alternative versions of anticipated reciprocity that eliminated the cooperative rule for the last period,¹⁰ but the results for these versions were essentially the same.

Tit-for-Tat Results

Estimating the TFT model, we found a maximized LL value of -8530.01 , which is a substantial gain from the basic rule learning model, but still less than rule learning with a pure cooperative rule. Furthermore, the estimates for the TFT parameters are $(c, \gamma, \delta_{\text{TFT}}) = (0.04, 5.03, 0.047)$, respectively. Given these parameters, $\psi(t+1; f_c(t), c)$ is a sharp step function at 0.04, implying that as long as one person out of 25 in the population chose the cooperative action in the period t , the TFT rule will cooperate in period $t+1$. This behavior is not much different from unconditional cooperation.

Since our specification of $\psi(t+1; f_c(t), c)$ with $c > 0$ excludes pure cooperation, we decided to test another specification that would admit pure cooperation. This was accomplished simply by letting c take on negative values: $c \in [-0.01, 1)$. Not surprising, the maximum LL was achieved for parameter values that mimicked a pure cooperation rule. Therefore, we cannot reject the hypothesis that the TFT rule adds nothing to explaining the data over a pure cooperation rule.

⁹ At the suggestion of a referee, we considered an encompassing model with a separate transference parameter for the anticipated reciprocity variable; however, we found no improvement in the maximized LL whatsoever and therefore cannot reject the hypothesis of a single transference parameter.

¹⁰ For the last period of each run in another version.

DISCUSSION

We have demonstrated how the framework of rule learning can accommodate behavioral rules like aspiration-based experimentation and reciprocity-based cooperation. In the data set considered consisting of 50 runs of 12 periods each involving 625 subjects, we found no evidence of aspiration-based experimentation or reciprocity-based cooperation. We did, however, find weak evidence for a simple experimentation by random belief rule that is reinforced by its actual expected payoff, and strong evidence for a simple cooperation rule that is also reinforced by its actual expected payoff like all other rules in the rule learning model. Further, we found no evidence for a tit-for-tat rule.

However, our conclusions regarding a simple cooperation rule must be softened. While the maximized LL dramatically increases with the addition of a simple cooperation rule, this gain comes largely by moving the mean prediction for the first period of Game 16 closer to the separatrix. Since the mean starting point for the dynamics is so close to the separatrix, random fluctuations around the mean generate starting points on both sides of the separatrix, which lead to final outcomes being a fairly even split between the two equilibria. However, simulated paths rarely exhibit the reversal and separatrix crossing pattern (like Fig. 1) seen in the data. Since the leading theoretical explanations of such behavior (aspiration-based experimentation and anticipated reciprocity) fail to be significant on this data set, we still have a behavioral pattern without an adequate explanation.

REFERENCES

- Anderson, J. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.
- Axelrod, R., & Hamilton, W. (1981). The evolution of cooperation. *Science*, **211**, 1390–1396.
- Binmore, K., & Samuelson, L. (1997). Muddling through: noisy equilibrium selection. *Journal of Economic Theory*, **74**, 235–265.
- Brandts, J., & Charness, G. (1999). *Punishment and reward in a cheap-talk game*. Unpublished manuscript.
- Camerer, C., & Ho, T. (1998). EWA learning in coordination games: probability rules, heterogeneity, and time-variation. *Journal of Mathematical Psychology*, **42**, 305–326.
- Camerer, C., & Ho, T. (1999a). EWA learning in games: preliminary estimates from weak-link games. In D. Budescu, I. Erev, & R. Zwick, (Eds.), *Games and human behavior: Essays in honor of Amnon Rapoport*. Hillsdale, NJ: Erlbaum.
- Camerer, C., & Ho, T. (1999b). Experience-weighted attraction learning in normal form games. *Econometrica*, **67**, 827–874.
- Camerer, C., Ho, T., & Chong, J. (2000). *Sophisticated EWA learning and strategic teaching in repeated games*. (Working paper 00-005). Wharton: Univ. of Pennsylvania.
- Cheung, Y-W., & Friedman, D. (1997). Individual learning in normal form games: some laboratory results. *Games and Economic Behavior*, **19**, 46–76.
- Cheung, Y-W., & Friedman, D. (1998). Comparison of learning and replicator dynamics using experimental data. *Journal of Economic Behavior and Organization*, **35**, 263–280.
- Churchland, P., & Sejnowski, T. (1992). *The computational brain*. Cambridge, UK: MIT Press.
- Cooper, R., Dejong, D., Forsythe, R., & Ross, T. (1993). Forward induction in the battle-of-the-sexes games. *American Economic Review*, **83**, 1303–1316.

- Dufwenberg, M., & Kirchsteiger, G. (1998). *A theory of sequential reciprocity*. (CentER Discussion Paper 9837). The Netherlands: Tilburg University.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Engle-Warnick, J., & Slonim, R. L. (2001). *The fragility and robustness of trust in repeated games*. (working paper). Case Western Reserve University.
- Erev, I., Bereby-Meyer, Y., & Roth, A. (1999). The effect of adding a constant to all payoffs: experimental investigation, and implications for reinforcement learning models. *Journal of Economic Behavior and Organization*, *39*, 111–128.
- Erev, I., & Roth, A. (1998). Predicting how people play games: reinforcement learning in experimental games with unique mixed strategy equilibria. *American Economic Review*, *88*, 848–881.
- Fehr, E., & Gächter, S. (1998). Reciprocity and economics: the economic implications of homo reciprocans. *European Economic Review*, *42*, 845–859.
- Fudenberg, D., & Levine, D. (1998). *The theory of learning in games*. Cambridge, MA: MIT Press.
- Gigerenzer, G., Todd, P., & the ABC Research Group, (1999). *Simple heuristics that make us smart*. Oxford: Oxford Univ. Press.
- Gneezy, U., Guth, W., & Verboven, F. (2000). Presents or investments? An experimental analysis. *Journal of Economic Psychology*, *21*, 481–493.
- Goffe, W., Ferrier, G., & Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, *60*, 65–99.
- Harsanyi, J., & Selten, R. (1988). *A general theory of equilibrium selection in Games*. Cambridge, MA: MIT Press.
- Haruvy, E., & Stahl, D. (1999). *Empirical tests of equilibrium selection based on player heterogeneity*, mimeograph.
- Ho, T-H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best-response in experimental *p*-beauty contests. *American Economic Review*, *88*, 947–969.
- Hoffman, E., McCabe, K., & Smith, V. (1998). Behavioral foundations of reciprocity: experimental economics and evolutionary psychology. *Economic Inquiry*, *36*, 335–352.
- Kreps, D., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoner's dilemma. *Journal of Economic Theory*, *27*, 245–252.
- Nagel, R. (1995). Unraveling in guessing games: an experimental study. *American Economic Review*, *85*, 1313–1326.
- Nelder, J., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*, 308–313.
- Newell, A., Shaw, J., & Simon, H. (1958). Elements of a theory of human problem solving. *Psychological Review*, *65*, 151–166.
- Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic*, *83*, 1281–1302.
- Roth, A., & Erev, I. (1995). Learning in extensive form games: experimental data and simple dynamic models in the intermediate term. *Games en Economic Behavior*, *8*, 164–212.
- Sarin, R., & Vahid, F. (1999). Payoff assessments without probabilities: a simple dynamic model of choice. *Games and Economic Behavior*, *28*, 294–309.
- Sen, A. (1977). Rational fools: a critique of the behavioral foundations of economic theory. *Journal of Philosophy and Public Affairs*, *6*, 317–344.
- Stahl, D. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior*, *16*, 303–330.
- Stahl, D. (1999). Evidence based rules and learning in symmetric normal form games. *International Journal of Game Theory*, *28*, 111–130.
- Stahl, D. (2000a). Rule learning in symmetric normal-form games: theory and evidence. *Games and Economic Behavior*, *32*, 105–138.

- Stahl, D. (2000b). *Action-reinforcement learning versus rule learning*. Department of Economics, University of Texas.
- Stahl, D. (2001). Population rule learning in symmetric normal-form games: theory and evidence. *Journal of Economic Behavior and Organization*, **45**, 19–35.
- Stahl, D., & Wilson, P. (1994). Experimental evidence of players' models of other players. *Journal of Economic Behavior and Organization*, **25**, 309–327.
- Stahl, D., & Wilson, P. (1995). On players models of other players: theory and experimental evidence. *Games and Economic Behavior*, **10**, 218–254.

Received: December 19, 2000; revised: October 2, 2001; published online: June 13, 2002