ELSEVIER

Journal of Economic Behavior & Organization
Vol. 1304 (2001) 1–17

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

Population rule learning in symmetric normal-form games: theory and evidence

Dale O. Stahl*

Department of Economics, University of Texas, Austin, TX 78712, USA

Received 22 January 1999; received in revised form 31 August 1999; accepted 31 August 1999

Abstract

A model of population rule learning is formulated and estimated using experimental data. When predicting the population distribution of choices and accounting for the number of parameters, the population rule learning model is much better than aggregation of individually estimated rule learning models. Further, rule learning is a statistically significant and important phenomena even when focusing on population statistics, and is much better than one-rule learning dynamics.

JEL classification: C15; C52; C72

Keywords: Rules; Learning; Games; Experimental; Testing

1. Introduction

Recent learning research in one-shot games can be divided into two domains: (i) population learning or evolutionary dynamics as typified by replicator dynamics,[1] and (ii) individual learning.[2,3] The first domain focuses on how the population distribution of play changes over time, while the second domain focuses on how an individual's behavior changes over time.

Individualistic models are needed for investigating the nature and characteristics of individual learning patterns and for assessing the amount of diversity in the population. Further,

* Tel.: +1-512-475-8541; fax: +1-512-571-3510.
E-mail address: stahl@eco.utexas.edu (D.O. Stahl).

[1] For example, Hofbauer and Sigmund (1988), Van Huyck et al. (1994), and Cheung and Friedman (1997).

[2] For example, Cheung and Friedman (1997), Crawford (1994), Mookherjee and Sopher (1994), Cooper and Feltovich (1996), Camerer and Ho (1997, 1999), Rapoport et al. (1997), and Stahl (1996, 1997a,b, 1999).

[3] Studies that examine both include Bush and Mosteller (1955), Friedman et al. (1998), and Roth and Erev (1995).

0167-2681/01/$ – see front matter
PII: S0167-2681(00)00169-4

if one wants to construct a cognitive theory of individual behavior in games, then individualistic models are essential because individual details could be masked in population statistics.

From a decision-theoretic framework, however, for one-shot games it is necessary and sufficient for a player to have a belief about the other players' actions, and when the other players are randomly drawn from a population of potential players such a belief is equivalent to a forecast of the population distribution of other players' actions. It is neither necessary nor sufficient to know anything about a single individual's learning dynamics, since one's actual opponents are random draws from a population. For example, to know which side of the road to drive on in the US, I do not need to know any specific history about the driver approaching me on the highway; I only need to know that in the US all sober drivers stay on the right side of the road.

Ideally, as in general equilibrium economics, one would like a theory of individual learning that aggregates up to a theory of population learning. However, we will encounter similar difficulties in finding aggregation theorems with reasonable assumptions. Of course, we can estimate individualistic models and then aggregate. But there is only so much information in any given dataset. If it is used to estimate a multitude of parameters of individualistic models, it does not follow that the prediction following aggregation is better than a prediction from a population (or representative agent) model with far fewer parameters. We will address this pertinent empirical question.

We focus on the class of rule learning models of Stahl (1996, 1997a,b, 1999) (hereafter S96, S97a,b, and S99). This is a rich class of learning models that encompasses action reinforcement (Roth and Erev, 1995; Erev and Roth, 1998), fictitious play (Brown, 1951), and belief updating (Mookherjee and Sopher, 1994; Camerer and Ho, 1997, 1999). [4] Briefly, a "rule" is a mapping from the game and history of play to a mixed strategy. For example, a noisy best response to the recent past is a Cournot-like rule that describes much of the behavior observed in experiments. Iterating once more we have a "level-2" rule that is a noisy best response to the best response to the recent past.

Complicating the econometric estimation of rule learning models is the fact that the rule used by an individual is not directly observable — only the action taken is observable — and in any model with properly specified error structures all rules will have full support on the available undominated actions. In an individualistic model of rule learning (S97b), the posterior probability of the rule conditional on the history was computed, but the computational complexity necessitated the use of precarious approximations. This problem can be potentially avoided by a population learning model because the experience of many individuals using and evaluating different rules gets merged into the population experience, so in essence it is as if the population evaluates all the rules.

In Section 2 we review the individual rule learning model of S97a, spell out aggregation of that model, and develop a population version of rule learning. Section 3 describes the experimental design and data, and Section 4 describes the econometric specification and computational issues. Section 5 presents the results, and Section 6 discusses our findings.

[4] While Camerer and Ho consider both reinforcement learning and belief learning, they formulate a single hybrid rule that combines these two aspects rather than allowing both types of rules to exist simultaneously in the population. In contrast, our rule learning model allows for many rules to exist simultaneously in the population and in the minds of players.

2. Theory

We begin with a description of the game environment and then present the theory and operational specifics of the model to be tested in this paper. For a more in-depth description of the general theory of rule learning, see S97a and S99.

2.1. 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, $A \equiv \{1, \ldots, J\}$ is the set of actions available to each player, and 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 $p^0 \equiv (1/J, \ldots, 1/J')$ denote the uniform distribution over actions A.

Let I denote the population of individuals from which the players are randomly drawn. We will use a superscript i to denote a particular individual player from I. We focus on single population situations in which each player is matched in every period with every other player in I; hence, the payoff relevant statistic for any given player is the probability distribution of the choices of the other players in the population, and this information is available to the players. [5] To this end, p^t will denote the empirical frequency of all players' actions in period t, and p^{it} will denote the empirical frequency of the actions of all players other than player i. The first period of play will be denoted by $t = 1$. It is also convenient to define $h^t \equiv \{p^0, \ldots, p^{t-1}\}$ as the history of all players' choices up to period t with the novelty that p^0 is substituted for the null history, and similarly $h^{it} \equiv \{p^{i0}, \ldots, p^{it-1}\}$ will denote the history observed by player i, where $p^{i0} \equiv p^0$ for all $i \in I$. Thus, the information available to player i at the beginning of period t is $\Omega^{it} \equiv (G, h^{it})$.

2.2. The general theory of rule learning

A behavioral rule is a mapping from information Ω^{it} to $\Delta(A)$, 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; $\rho(\Omega^{it})$ is the mixed strategy generated by rule ρ given information Ω^{it}.

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

$$\varphi^i(\rho, t) \equiv \frac{\exp(w^i(\rho, t))}{\int \exp(w^i(x, t))\,\mathrm{d}x}. \tag{1}$$

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

[5] See Crawford (1994) for an adaptive learning model for such situations.

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

Computing this integral is the major computational burden of this model.

The last element of the general theory 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:

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

where $g(\cdot)$ is the reinforcement function for rule ρ conditional on information $\Omega^{it+1} = (G, h^{it+1})$. The inertia parameter β_0 determines how much weight is given to the past versus new reinforcement information. It is natural to assume that $g(\rho, \Omega^{it+1})$ is proportional to the expected utility that rule ρ would have generated in period t:

$$g(\rho, \Omega^{it+1}) = \beta_1 \hat{p}(\Omega^{it}; \rho) U p^{it}, \tag{4}$$

where $\beta_1 > 0$ is a scaling parameter that converts expected utility units into log-propensity units. Then, for small β_0 and large β_1, past propensities would be quickly swamped by new performance evidence, whereas for large β_0 and small β_1, past propensities would persist despite new evidence.

Given a space of rules R and initial conditions $w^i(\cdot, 1)$, 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 $w^i(\cdot, 1)$.

An attractive feature of this rule learning 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. Moreover, the general model can include these constant rules, best-response rules and other rules.

2.3. Aggregation

Let μ be a probability measure on the population, I, of individuals. Then aggregation of choice probabilities would give

$$\hat{p}(t) \equiv \int_I \hat{p}^i(t)\, \mathrm{d}\mu = \int_I \int_R \rho(\Omega^{it})\, \mathrm{d}\varphi^i(\rho, t)\, \mathrm{d}\mu. \tag{5}$$

For a model of "population learning" not based on aggregation of individual behavior, we (i) use p^t, the distribution of choices of the whole population,[6] in lieu of the set $\{p^{it}\}_{i \in I}$; (ii) apply the law of motion to $w(\rho, t) = \int_I w^i(\rho, t)\, \mathrm{d}\mu$:

[6] We are implicitly assuming that the population is large enough so the influence of one member is negligible. Without this assumption, the personalized information of each player would be needed for econometric estimation, which would increase the computational complexity, make the state space for conditional probabilities grow exponentially over time, and render the model inapplicable to datasets with only aggregated choice information.

$$w(\rho, t+1) = \beta_0 w(\rho, t) + (1-\beta_0) g(\rho, \Omega^{t+1}), \quad \text{for } t \geq 1, \tag{3'}$$

and (iii) determine population rule probabilities, $\varphi(\cdot, t)$, by

$$\varphi(\rho, t) \equiv \frac{\exp(w(\rho, t))}{\int \exp(w(x, t))\, \mathrm{d}x}. \tag{1'}$$

However, note that because of the non-linearity of the logit function, Eq. (1′), the population rule probabilities, $\varphi(\rho, t)$, as defined by Eqs. (1′) and (3′), is not necessarily the same as predicted by aggregation of the individual rule probabilities: $\int_I \varphi^i(\rho, t)\, \mathrm{d}\mu$. In other words, using Eqs. (1′) and (3′) in our population learning model will introduce a potential specification error. Of course, if the population were homogeneous, then the population model using Eqs. (1′) and (3′) would be valid; however, homogeneity is soundly rejected by empirical evidence (S97a).

Offsetting this potential specification error is a more palatable interpretation of rule evaluation. Whereas Eq. (3) implicitly assumes that all rules are evaluated each period by every individual, Eq. (3′) can be interpreted as requiring only that every rule is evaluated by some individual and that the population as a whole looks like a representative individual who has evaluated all the rules. We will econometrically estimate this "population rule learning" model, and compare its predictive performance to aggregation of the individual rule learning model.

2.4. 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. We will use the family of "evidence-based" rules which was introduced in S97a,b and S99, as an extension of the Stahl and Wilson (1995) (hereafter SW95) 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 which has the most net favorable evidence based on the available information.

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 strategy, so for the first period of play by virtue of insufficient reason, the belief is that all strategies are equally likely. The expected utility payoff to each available action given the null model is $y_1(\Omega^{i1}) \equiv Up^0$. 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: $(1-\theta)p^0 + \theta p^{i1}$ for period 2 with $\theta \in [0, 1]$. Letting $q^{it}(\theta)$ denote the forecast for period t and defining $q^{i0}(\theta) \equiv p^0$, the following forecasting equation applies for all $t \geq 1$:

$$q^{it}(\theta) \equiv (1-\theta) q^{it-1}(\theta) + \theta p^{it-1}. \tag{6}$$

The expected utility payoff given this belief is $y_1(\Omega^{it}; \theta) \equiv Uq^{it}(\theta)$. We can interpret $y_{1j}(\Omega^{it}; \theta)$ as "level-1" evidence in favor of action j stemming from the null model and prior history h^{it}.

The second kind of evidence is based on the SW95 "level-2" player who believes all other players are level-1 players, and hence believes that the distribution of play will be $b(q^{it}(\theta))$, where $b(q^{it}(\theta)) \in \Delta(A)$ puts equal probability on all best responses to $q^{it}(\theta)$ and zero probability on all inferior responses. The expected utility conditional on this belief is $y_2(\Omega^{it}; \theta) \equiv Ub(q^{it}(\theta))$. We can interpret $y_{2j}(\Omega^{it}; \theta)$ as "level-2" evidence in favor of action j.

The third kind of evidence incorporates Nash equilibrium theory within the model. Letting p^{NE} denote a Nash equilibrium of G, $y^3 \equiv Up^{\mathrm{NE}}$ provides yet another kind of evidence on the available actions.

Finally, 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 perhaps some inertia as represented by $q^{it}(\theta)$. We then define $y_0(\Omega^{it}; \theta) = \ln[q^{it}(\theta)]$, so when entered into the logistic equation (1) as the log-propensity the resulting choice probabilities would be $q^{it}(\theta)$.

So far we have defined four kinds of evidence: $Y \equiv \{y_0, \ldots, y_3\}$. The next step is to weigh this evidence and specify a probabilistic choice function. Let $\nu^k \geq 0$ denote a scalar weight associated with evidence y_k. We define the weighted evidence vector:

$$\bar{y}(\Omega^{it}; \nu, \theta) \equiv Y(\Omega^{it}; \theta)\nu, \tag{7}$$

where $\nu \equiv (\nu_0, \ldots, \nu_3)'$.

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 which from his/her perspective has the greatest net favorable evidence. Hence, the probability of choosing action j is

$$\hat{p}_j(\Omega^{it}; \nu, \theta) \equiv \frac{\exp[\bar{y}_j(\Omega^{it}; \nu, \theta)]}{\sum_\ell \exp[\bar{y}_\ell(\Omega^{it}; \nu, \theta)]}. \tag{8}$$

Note that, given the five-dimensional parameter vector (ν, θ), Eq. (8) defines a mapping from Ω^{it} to $\Delta(A)$, and hence is a behavior rule as defined abstractly above. By putting zero weight on all but one rule, Eq. (8) defines an archetypal rule — one for each kind of evidence corresponding to the underlying model of other players. Eq. (7) generates the space of rules spanned by these archetypal rules.

For our population learning model, we simply drop the subscript i; that is, we use the population evidence rather than the evidence from any one player's individual experience.

2.5. Transference

Since this theory is about rules that use information about the game as input, we should be able to predict behavior in a temporal sequence that involves a variety of games. For instance, in our experiment one game is played for 15 periods and then another game is played for 15 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:

$$w^i(\rho, 16) = (1-\tau)w^i(\rho, 1) + \tau w^i(\rho, 15+), \tag{9}$$

where "15+" indicates the update after period 15 of the first run, and τ is the transference parameter. (For the population version, we simply drop the superscript i.) If $\tau = 0$, there is no transference, so period 16 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 16 of the first run (with no change of game). This specification extends the theory to any number of runs with different games without requiring additional parameters (beyond τ).

3. Experimental design and data

An experiment session consisted of two runs of 15 periods each. In the first run, one of the four games was played for 15 periods, and in the second run, another game was played for 15 periods. A "mean matching" protocol was used, i.e. in each period a participant was determined by his/her choice and the percentage distribution of the choices of all other participants: Up^t.

Four 5×5 symmetric games were selected to challenge the theory. The payoffs for the "row player" are shown in Fig. 1. The labels ne, b1, b2, wd, denote, respectively, the unique symmetric Nash equilibrium strategy, the best-response to uniform (level-1), the best response to the best response to uniform, and the choice of the SW95 "worldly" type which best responds to a convex combination of the Nash prior and a noisy level-2 prior. The fifth strategy was either a dominated strategy (dm), or the maximax strategy (mx). Payoffs are in probability units for a fixed prize of US$ 2.00 per game. The lotteries that determined final monetary payoffs were conducted following the completion of both runs.

Participants were seated at private computer terminals. Each game ("decision matrix") was presented on the computer screen. The participant made a choice of a pure strategy by clicking on a row of the matrix, which then became highlighted. In addition, the participant could enter a hypothesis about the choices of the other players, and cause the computer to calculate hypothetical earnings, which were then displayed on the screen. Following each period, each participant was shown the aggregate choices of the other participants for that period. At any time by clicking the Record button, the participants could view a Record screen with the history of the aggregate choices of the other participants for the entire run.

The experiment consisted of four sessions of 22, 23, 24 and 22 participants. The participants were predominately upper division undergraduate students and some graduate students attending the first and second 1995 summer sessions at the University of Texas. The average payment per participant was US$ 28.00 for a 2.5 h session. For a more complete description of the experiment and the instructions, see S97a.

Game I:

19	43	96	85	85	b1
28	62	88	74	24	ne
67	21	38	48	38	b2
40	58	0	15	92	wd
16	15	86	99	79	mx

Game II:

68	10	76	33	75	wd
73	4	59	0	8	dm
3	92	16	15	99	b2
86	54	25	41	6	ne
72	98	92	8	52	b1

Game III:

2	31	0	99	6	mx
6	10	97	40	24	b2
98	96	38	48	19	b1
42	40	80	51	48	wd
97	46	5	68	49	ne

Game IV:

22	79	35	56	75	wd
22	38	78	55	99	b1
27	58	1	11	0	dm
70	1	34	59	37	ne
56	84	60	23	2	b2

Fig. 1. The four games.

4. Econometric specification

The theoretical model put forth to explain these choices involves nine parameters: $\beta \equiv (\bar{\nu}_0, \bar{\nu}_1, \bar{\nu}_2, \bar{\nu}_3, \bar{\theta}, \sigma, \beta_0, \beta_1, \tau)$. The first five parameters $(\bar{\nu}_0, \ldots, \bar{\nu}_3, \bar{\theta})$ represent the mean of the participant's initial probability φ, and σ is the standard deviation of φ; β_0 and β_1 are the learning parameters of Eqs. (2) and (3); and τ is the transference parameter in Eq. (9) for the initial propensity of the second run.

Let $s^i \equiv (s^{i1}, \ldots, s^{i30}) \in \{1, 2, 3, 4, 5\}^{30}$ denote the choices of participant h for an experiment, and let n_j^t denote the number of participants who choose action j in period t. For computational issues, see Appendix A.

4.1. The individual rule learning likelihood function

Letting $\varphi^i(\nu, \theta, t|\beta^i)$ denote the probability that participant i uses rule (ν, θ) in period t (with information Ω^{it}) given the nine-parameter vector β^i, the resulting probability that i chose action j is

Table 1
Prediction of population choices

Session	AL	LA	AIC
S627	−530.26	−545.14	−348.24
S629	−894.43	−891.71	−401.44
S810	−819.29	−822.43	−407.72
S815	−817.25	−803.17	−406.16
Total	−3061.23	−3062.45	−1563.56

$$p_j^{it}(\beta^i) = \int_R \hat{p}_j(\Omega^{it}; \nu, \theta)\varphi^i(\nu, \theta, t|\beta^i)\,\mathrm{d}(\nu, \theta). \tag{10}$$

Then the joint probability of s^i conditional on β^i is

$$P^i(\beta^i) \equiv \prod_{t=1}^{30} p_{s^{it}}^{it}(\beta^i). \tag{11}$$

S97a found the parameter values $\hat{\beta}^i$ that maximized $\log[P^i(\beta^i)]$ for each participant. Given N participants, parameter estimates $\underline{\beta} \equiv (\beta^1, \ldots, \beta^N)$, and letting $\mu = 1/N$ in Eq. (5), the predicted population distribution of choice j in period t is

$$p_j^t(\underline{\beta}) \equiv \sum_i \frac{p_j^{it}(\beta^i)}{N}. \tag{12}$$

Given an urn filled with the participants, $p_j^t(\underline{\beta})$ is the probability that a randomly drawn participant will choose action j (assuming that the β^i of any participant is not observable), and hence $p_j^t(\underline{\beta})$ is also the ex ante expected population distribution from random draws from the urn with replacement. Then the predicted log-likelihood of the aggregated choices $\{n^t, t = 1, \ldots, 30\}$ is

$$\mathrm{AL}(\underline{\beta}) \equiv \sum_t \sum_j n_j^t \log[p_j^t(\underline{\beta})]. \tag{13}$$

It is important to recognize that $\sum_i \log[P^i(\beta^i)] \neq \mathrm{AL}(\underline{\beta})$.[7] Therefore, the parameter values $\hat{\beta}^i$ that maximize the former (the aggregate log-likelihood of the individual choices) do not maximize the log-likelihood of the aggregated choices. The values of $\mathrm{AL}(\underline{\hat{\beta}})$ based on the parameter estimates of S97a are given in Table 1.

[7] To see this, note that the former expression is equivalent to summing the log-likelihood of every individual's choice for every period, while the latter first sums the choice probabilities over all individuals period by period, and then sums the logarithms of those aggregated choice probabilities. These two expressions yield identical values if and only if $\beta^i = \beta$ for all i.

4.2. The population rule learning likelihood function

The rule propensities and law of motion is given by Section 2.2, with Eqs. (1′) and (3′) for the population distributions. A single-parameter vector β applies to the population, yielding population choice probabilities:

$$p_j^t(\beta) = \int_R \hat{p}_j(\Omega^t; \nu, \theta)\varphi(\nu, \theta, t|\beta)\,\mathrm{d}(\nu, \theta). \tag{14}$$

Then the log of the joint probability of the aggregate data conditional on β is

$$\mathrm{LA}(\beta) \equiv \sum_t \sum_j n_j^t \log[p_j^t(\beta)], \tag{15}$$

analogous to Eq. (13).

5. Results

We estimated a homogeneous model (one nine-parameter β vector for all four experimental sessions), and we also estimated separate models for each session. The maximized log-likelihood value for the homogeneous model is -3097.01, while the sum of the four maximized log-likelihood values for the session models is -3062.45. Twice the difference is distributed chi-square with 27 (3×9) degrees of freedom and has a P-value of 10^{-5}. Therefore, we reject the hypothesis that the four sessions come from the same distribution. [8] Henceforth, we will report on only disaggregated session-by-session results.

5.1. Population model versus aggregation of individual models

The session-by-session maximized log-likelihood values, LA(β), of the population rule learning model are given in the third column of Table 1 next to the computed values of AL($\underline{\beta}$) based on S97a. It is immediately apparent that there is little difference in the log-likelihood values both by session and aggregated over all sessions. Therefore, the slight overall improvement in the aggregated log-likelihood provided by aggregation does not appear to be worth the tremendous increase in the number of parameters: $(9 \times 91) - (9 \times 4) = 783$. One method of comparing log-likelihood values of non-nested models with different number of parameters is the Akaike's information criteria, which is twice the log-likelihood difference less twice the difference in the number of parameters. This measure is given in the fourth column of Table 1. Clearly according to the AIC, the population model is far superior. Modifications of the AIC that have been suggested in the literature (Bozdogan, 1987) only serve to enhance the effect of the difference in the number of parameters. Therefore, we have the following result.

[8] The variance in the parameter estimates across sessions is consistent with the estimated variance in individual parameter estimates found in S97a.

Table 2
Coefficient estimates of population model

Parameter	S627	S629	S810	S815
$\bar{\nu}_0$	0.054^{+}	0.026	0.328**	0.529**
$\bar{\nu}_1$	0.096**	0.053**	0.106**	0.065**
$\bar{\nu}_2$	0.000	0.001	0.000	0.000
$\bar{\nu}_3$	0.000	0.000	0.003	0.001
$\bar{\theta}$	0.302**	0.733**	0.250**	0.468**
σ	0.671	1.04	0.306	0.704
β_0[a]	0.878**	0.938**	0.814**	0.893
β_1	0.060**	0.007	0.021**	0.011
τ	0.626**	1.00**	1.00**	0.000

[a] β_0 is tested relative to 1.0.
$^{+}$ Significant at 10% level.
** Significant at 1% level.

Result 1. Taking account of the difference in the number of parameters, the population rule learning model is superior to aggregation of the individual rule learning model. [9]

The coefficient estimates of the population rule learning model are given in Table 2. It is noteworthy that the initial weights on level-2 evidence ($\bar{\nu}_2$) and Nash evidence ($\bar{\nu}_3$) are insignificant from zero, whereas the initial weight on level-1 evidence is highly significant. [10]

5.2. Rule learning hypotheses

Two parameters are critical to rule learning: β_0 and β_1. If $\beta_0 = 1$, then by virtue of law of motion, Eq. (3′), the population distribution of rule propensities would be constant for all periods. The test reported in Table 2 for β_0 is with respect to $\beta_0 = 1$, and this null hypothesis is rejected for three out of the four experimental sessions. Aggregating all sessions, the null hypothesis has a P-value less than 10^{-7}. Therefore, we have the following result.

Result 2. We strongly reject the hypothesis of constant population rule propensities.

While rule propensities apparently change over time, they respond to the performance evaluation function if and only if β_1 is significantly positive. From Table 2, it can be seen that the null hypothesis of $\beta_1 = 0$ is rejected for two of the four sessions. Aggregating all

[9] Cheung and Friedman (1997) find that a representative agent model does not fit the data as well as individual agent models. Specifically, they reject the hypothesis that $\beta^i = \beta$ for all i when maximizing the sum of individual log-likelihoods: $\sum_i \log[P^i(\beta^i)]$. The same hypothesis is also rejected on our dataset (S97a). These tests presume that we can identify an individual participant i with a parameter vector β^i, which we can do ex post. However, when considering the predictive power of a model, the best we can hope for ex ante is to know the distribution of the parameter vectors in the participant population. Hence, the appropriate ex ante log-likelihood measure is that given by Eq. (13), which is derived from the urn model. Accordingly, Result 1 says that the population diversity that is implicit in the population rule learning model is as good as the diversity captured by an urn containing unobservable individual parameter estimates.

[10] Note that σ was not tested, since the log-propensities are unbounded for $\sigma = 0$.

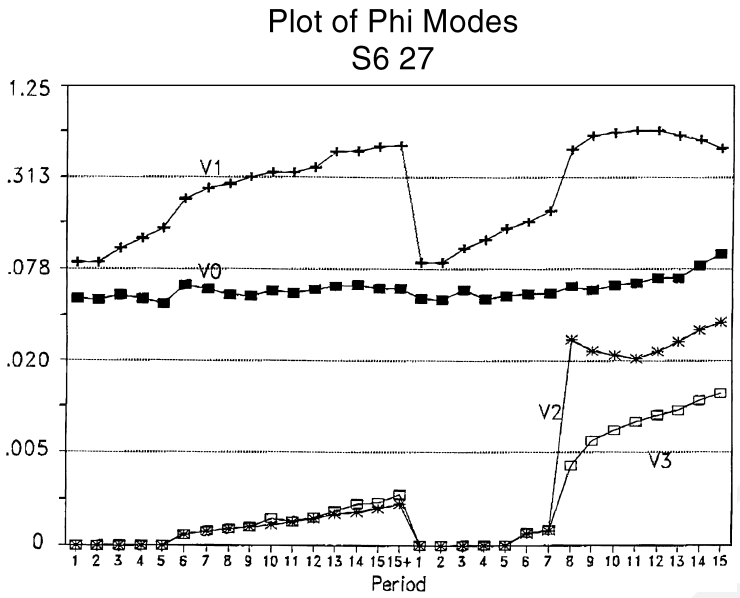

Fig. 2. Plot of φ mode for S627.

sessions, the null hypothesis has a P-value less than 10^{-7}. Therefore, we have the following result.

Result 3. We reject the hypothesis of no rule learning.

If we had only two dimensions or only a small number of rules, then we could easily present a potentially revealing plot of rule propensities φ over time. However, with five dimensions, it is a challenge to present a picture of how the probability distribution over rules (φ) changes over time. For each period we identified the "dominant mode" of φ as follows. We found all the rules (ν, θ) for which $\varphi(\nu, \theta, t)$ was within 50% of the maximum φ value for that period, and computed the average, $\int(\nu, \theta)\, \mathrm{d}\varphi(\nu, \theta, t) / \int \underline{1}\, \mathrm{d}\varphi(\nu, \theta, t)$, over this neighborhood of the dominant mode; call this $(V_0^t, V_1^t, V_2^t, V_3^t, \Theta^t)$ for period t. Figs. 2 and 3 display these mean evidence weights for the dominant mode as a function of time for the two sessions for which β_1 was statistically significant. For S627, the transference parameter (τ) is essentially zero, so the V_k values revert to the initial values at the beginning of the second run; while for S810, $\tau \approx 1$, so the V_k values are constant between runs. Note that an increase of one unit on the log scale means a four-fold increase in the weight on the corresponding evidence. Hence, these figures reveal substantial changes in φ due to rule learning, especially for the level-2 and Nash rules. In Fig. 2, the weight on level-l evidence increases throughout the first run and during the first half of the second run, when the weight on level-2 and Nash evidences increases dramatically. In Fig. 3, little happens during the first 10 periods, but then the weights on level-2 and Nash evidences steadily increase.

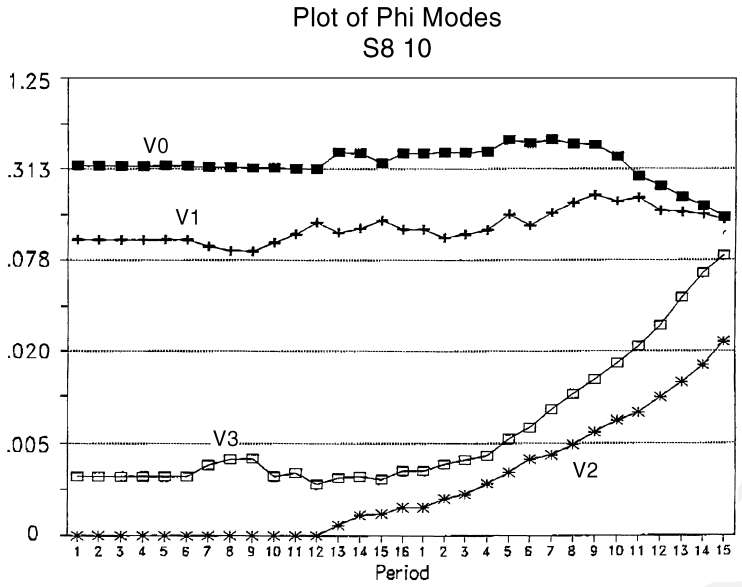

Fig. 3. Plot of φ mode for S810.

5.3. *Nash and Cournot hypotheses*

As a benchmark for the population rule learning model, we can consider the Nash equilibrium model. Of course, in its pure form, it is incompatible with the data because participants often make non-Nash choices. It is more interesting to consider the Nash model extended to include errors. Observe that by setting $\nu_k = 0, \forall k \neq 3$, we have a Nash-based probabilistic choice function, with the interpretation of ν_3 as the precision of the population's expected utility calculation or as the inverse of the variance of the population's idiosyncratic considerations. The hypothesis that the population makes its choices according to this error-prone Nash model is nested within our full population model as a seven-parameter restriction. For each session, we found the $(\bar{\nu}_3, \bar{\theta})$ values that maximized the log-likelihood of the population choices. The sum over all sessions of these maximized L values was -4087.78. Compared with the totally random prediction (-4393.75), this is a significant improvement ($P < 10^{-126}$). However, the full population model (-3062.44) is a very significant improvement over this Nash model ($P < 10^{-417}$). In other words, even after adjusting for the large number of parameters, the population rule learning model is astronomically more likely to have generated the data than the Nash-based model. (For an enhanced Nash model with learning which is also rejected, see Appendix A.)

Result 4. We strongly reject the implicit restrictions of the Nash model.

So-called Cournot dynamics have been popular because of their simplicity and explanatory power (e.g. Van Huyck et al., 1994; Cheung and Friedman, 1997; Friedman et al.,

1995). In our context, Cournot dynamics is equivalent to zero weight on all evidence except $y_1(\Omega^t, \theta)$, and no rule learning. Thus, the reduced model would have only two parameters $(\bar{\nu}_1, \bar{\theta})$. Maximizing the log-likelihood function with respect to these two parameters for each session and summing over all four sessions, the aggregated log-likelihood decreases to -3321.66. Compared to the no-rule-learning model, twice the difference is distributed chi-square with 12 degrees of freedom and has a P-value less than 10^{-91}. Compared to the full population model, twice the difference is distributed chi-square with 28 degrees of freedom and also has a P-value less than 10^{-91}. Thus, we can strongly reject the Cournot model in favor of both the no-rule-learning model (but other rules present) and the full population model. (For an enhanced Cournot model with learning which is also rejected, see Appendix A.)

Result 5. We strongly reject "Cournot dynamics". [11]

6. Discussion

We draw two main conclusions from this study. First, when predicting the population distribution of choices, our population rule learning model, with its implicit population diversity, is as good as, and accounting for the number of parameters is much better than, aggregation of individually estimated rule learning models. Hence, for the purposes of developing a descriptive theory of population choices and a prescriptive theory of play, we should focus on population rule learning models. For other learning models that cannot represent population diversity, the adequacy of population models remains an open question.

Second, rule learning is a statistically significant and important phenomena even when focusing on population statistics, and is much better than one-rule learning dynamics such as "Cournot dynamics" and Nash equilibrium learning.

Although we reported only session-by-session results, we performed all the tests of Section 5 on a model with a single set of parameters for all session and found similar results. In on going research, we are investigating an expanded rule learning model with constant rules as well as the evidence-based rules, encompassing the major competing learning theories from action reinforcement to sophisticated rule learning, thereby permitting nested hypothesis testing. We will also explore parsimonious ways to capture heterogeneity within the population — such as mixture models.

Uncited reference

Cheung and Friedman (1998).

[11] Since fictitious play can be represented as a variation on Cournot dynamics with the θ parameter of Eq. (6) being time-dependent (specifically, $\theta = 1/t$), and since θ is allowed to vary over periods in our rule learning model, we can also reject fictitious play.

Acknowledgements

Partial funding of this research was provided by Grant nos. SBR-9410501 and SBR-9631389 from the National Science Foundation, but nothing herein reflects their views. Ray Battalio assisted in the design of the computer interface, and Ernan Haruvy provided research assistance. However, all errors and omissions are the sole responsibility of the author.

Appendix A

A.1. Computational methods and issues

To perform the integration required by Eqs. (10) and (14), we imposed a finite grid on the (ν, θ)-space.[12] As in S97a, we used the following logarithmic grid: $\nu_k \in \{5/4^{j-1},\ j = 1, 6\} \cup \{0\}$, and $\theta \in \{j/6,\ j = 0, 6\}$. We also included the mean parameter $(\bar{\nu}_0, \ldots, \bar{\nu}_3, \bar{\theta})$ as a variable point of the grid. Thus, the grid consisted of $7^5 + 1 = 16{,}808$ points.[13] Consistent with the grid scaling, we confined the $\bar{\nu}_k$ estimates to the $[0, 5]$ interval and the $\bar{\theta}$ estimate to the $[0, 1]$ interval.

The initial log-propensity function was specified as

$$w(\nu, \theta, 1) = -\frac{0.5\|(\nu, \theta) - (\bar{\nu}, \bar{\theta})\|^2}{\sigma^2}, \tag{A.1}$$

where the distance $\|(\nu, \theta) - (\bar{\nu}, \bar{\theta})\|$ was measured on this logarithmic scale. That is, each ν_k was converted into a grid number, $1 + \ln(5/\nu_k)/\ln(4)$, and θ was assigned grid value $1 + 6\theta$; then the distance between the five-dimensional vector of grid numbers was computed and assigned to $\|(\nu, \theta) - (\bar{\nu}, \bar{\theta})\|$.

To find a β^i that maximizes $\log[P^i(\beta^i)]$, we used a simulated annealing algorithm (Goffe et al., 1994) for 13,502 function evaluations at high (but declining) temperatures, and then fed the result into the Nelder and Mead (1965) algorithm on a Cray J90. We found the simulated annealing algorithm to be effective in exploring the parameter space, but very slow to converge once it settled in on a local maximum, while the latter algorithm converged much faster locally. A typical estimation of the nine parameters for 91 participants took about 200 J90 cpu hours. We used the same method to find a β that maximizes $\mathrm{LA}(\beta)$.

A.2. A Nash model with learning

It may be objected that the Nash model of the text is inadequate because it does not permit the population to learn the Nash equilibrium over time, even though behavior converged to the Nash equilibrium in only two of the eight runs. Nonetheless, it might be informative to

[12] Since a priori we do not know whether the distribution φ in the second period is single- or multimodal, statistical integration techniques are not appropriate.

[13] Subsequent investigations with this dataset and a different grid produced qualitatively similar results, demonstrating that the results reported here are not an artifact of the grid.

consider the non-nested population model in which only the dimension of Nash-evidence rules are present and the population can learn how much weight to put on the Nash prior (or equivalently, can improve the accuracy of its expected utility calculations). We suppose that Eq. (3) applies, but only with respect to ν_3. There are five parameters of this "Nash-learning" model: σ, $\bar{\nu}_3$, β_0, β_1, and τ. The aggregated log-likelihood for this model is -3903.28, which is a substantial improvement over the above two-parameter Nash model, but still astronomically worse than the full population model. We can view this Nash-learning model as nested in an encompassing model in which there are two sets of (σ, β_0, β_1, τ) parameters, one for the ν_3 dimension and one for the (ν_0, ν_1, ν_2) dimensions combined. The five-parameter Nash-learning model is equivalent to restricting the second set of (σ, β_0, β_1, τ) parameters to be (0, 1, 0, $\cdot$), in addition to the restriction that $\nu_k = 0$ for $k \neq 3$. The 13-parameter encompassing model was not estimated, but it must have a log-likelihood at least as large as our full population model (-3062.45) since our full model is equivalent to restricting the two sets of (σ, β_0, β_1, τ) parameters to be the same. Then, twice the difference would be distributed chi-square with 32 ($(13-5)\times 4$) degrees of freedom, and is at least as great as $2 \times (3903.28 - 3062.45)$, and therefore has a P-value less than 10^{-333}.

A.3. A Cournot model with learning

It might be objected that the Cournot model of the text is too simplistic in that it does not permit learning of the θ parameter of the forecasting equation nor learning how much weight to put on the y_1 evidence. To answer this criticism, we estimated a non-nested model in which only the y_1-evidence dimension of rules are present, and the population can adjust θ and ν_1 in response to reinforcement evidence according to Eq. (3). There are six parameters in this "enhanced Cournot" model: $\bar{\nu}_1$, $\bar{\theta}$, σ, β_0, β_1, and τ. The aggregated log-likelihood for this enhanced Cournot model is -3204.28, which is a substantial improvement over the above two-parameter Cournot model, but still astronomically worse than the full population model. We can view this enhanced model as nested in an encompassing model in which there are two sets of (σ, β_0, β_1, τ) parameters, one for the ν_1 dimension and one for the (ν_0, ν_2, ν_3) dimensions combined. Our six-parameter enhanced Cournot model is equivalent to restricting the second set of (σ, β_0, β_1, τ) parameters to be (0, 1, 0, $\cdot$), in addition to the restriction that $\nu_k = 0$ for $k \neq 1$. The 13-parameter encompassing model was not estimated, but it must have a log-likelihood at least as large as our full model (-3062.45) since our full model is equivalent to restricting the two sets of (σ, β_0, β_1, τ) parameters to be the same. Then, twice the difference would be distributed chi-square with ($(13-6)\times 4$) 28 degrees of freedom, and is at least as great as $2 \times (3204.68 - 3062.45)$, and therefore has a P-value less than 10^{-43}.

References

Bozdogan, H., 1987. Model selection and Akaike's information criterion: the general theory and its analytical extensions. Psychometrika 52, 345–370.

Brown, G., 1951. Iterative solution of games by fictitious play. In: Koopmans, T. (Ed.), Activity Analysis of Production and Allocation. Wiley, New York, pp. 374–376.

Bush, R., Mosteller, F., 1955. Stochastic Models of Learning. Wiley, New York.

Camerer, C., Ho, T., 1997. EWA learning in games: preliminary estimates from weak-link games. In: Hogarth, R. (Ed.), Games and Human Behavior: Essays in Honor of Amnon Rapoport.
Camerer, C., Ho, T., 1999. Experience-weighted attraction learning in normal-form games, Econometrica, in press.
Cheung, Y.-W., Friedman, D., 1997. Learning in evolutionary games: some laboratory results. Games and Economic Behavior 19, 46–76.
Cheung, Y.-W., Friedman, D., 1998. A comparison of learning and replicator dynamics using experimental data. Journal of Economic Behavior and Organization 35, 263–280.
Cooper, D., Feltovich, N., 1996. Reinforcement-based learning vs. Bayesian learning: comparison. Mimeo, University of Pittsburgh.
Crawford, V., 1994. Adaptive dynamics in coordination games. Econometrica 63, 103–143.
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.
Friedman, D., Massaro, D., Cohen, M., 1995. A comparison of learning models. Journal of Mathematical Psychology 39, 164–178.
Goffe, W., Ferrier, G., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. Journal of Econometrics 60, 65–99.
Hofbauer, J., Sigmund, K., 1988. The Theory of Evolution and Dynamical Systems. Cambridge University Press, Cambridge.
Mookherjee, D., Sopher, B., 1994. Learning behavior in an experimental matching pennies game. Games and Economic Behavior 7, 62–91.
Nelder, J., Mead, R., 1965. A simplex method for function minimization. Computer Journal 7, 308–313.
Rapoport, A., Erev, I., Abraham, E., Olson, D., 1997. Randomization and adaptive learning in a simplified poker game. Organizational Behavior and Human Decision Processes 69, 31–49.
Roth, A., Erev, I., 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. Games and Economic Behavior 8, 164–212.
Stahl, D., 1996. Boundedly rational rule learning in a guessing game. Games and Economic Behavior 16, 303–330.
Stahl, D., 1997a. Rule learning in symmetric normal-form games: theory and evidence. Games and Economic Behavior, in press.
Stahl, D., 1997b. Local rule learning in symmetric normal-form games: theory and evidence. Mimeo.
Stahl, D., 1999. Evidence-based rules and learning in symmetric normal-form frames. International Journal of Game Theory 28, 111–130.
Stahl, D., Wilson, P., 1995. On players' models of other players: theory and experimental evidence. Games and Economic Behavior 10, 218–254.
Van Huyck, J., Cook, J., Battalio, R., 1994. Selection dynamics, asymptotic stability, and adaptive behavior. Journal of Political Economy 102, 975–1005.