Sophisticated Learning and Learning Sophistication

by

Dale O. Stahl

Malcolm Forsman Centennial Professor

Department of Economics

University of Texas

stahl@eco.utexas.edu

March 2003

ABSTRACT

We test the population rule learning model for symmetric normal-form games, and strongly reject: (i) no rule learning, (ii) no diversity, and (iii) no sophisticated evidence. Further, trembles and herd behavior decline and level-2 behavior increases over time.

*Partial funding of this research was provided by grant SBR-9410501 and SBR-9631389 from the National Science Foundation and the Malcolm Forsman Centennial Professorhip, but nothing herein reflects their views. I am especially indebted to Ernan Haruvy who helped design and run the experiments and provided invaluable intellectual stimulation. All errors and omissions are my sole responsibility.

1. Introduction.

Are human subjects in game theory experiments as simple minded as most learning theories assume they are?[1] At one extreme, the players in

[2]E.g. Selten (1990, 1991), Kandori, *et al* (1993), Mookherjee and Sopher (1994, 1997), Friedman, *et al* (1995), Roth and Erev (1995), Anderson, *et al* (1997), Camerer and Ho (1997, 1999), Cheung and Friedman (1997, 1998), and Erev and Roth (1998), to mention only a few.

evolutionary game theory could just as well be thoughtless amoeba; learning is a byproduct of natural selection, not a conscious internal process in the players' minds. Most of the learning dynamics ignore much of the information that human players might be assumed to have available: e.q. the history of play and knowledge of the game. Such information would enable them to compute (a) hypothetical payoffs from actions they could have chosen but did not, (b) the best reply to the recent past, and (c) iterates of the best reply mapping. In this paper, we call learning "sophisticated" only if the dynamics admit this richer information, especially the iterates of the best reply mapping. Whether or not human players incorporate such information and do these computations is an empirical question and should not be assumed away by the model.

With few exceptions, learning models also restrict attention to the actions as the objects of the learning process. An obvious shortcoming of this view is that nothing can be learned that the player can transfer to a new but similar situation. For example, the population may converge to action A in first game, but in the next game action A might be dominated. Hence, when the game changes, the learning model must be revised and reinitialized. Yet, researchers who are university professors know very well that what we teach (and hopefully students learn) are ways of thinking about problems: high level algorithms for recognizing essential features and solving problems.

In this spirit, the rule learning models of Stahl (1996, 1999a, 2000, 2001) hypothesize a rich space of *behavioral rules* which players can learn based on performance feedback. These rules span several levels of sophistication (level-n bounded rationality, as well as herd behavior and Nash behavior). We call this rule learning "learning sophistication". There is growing evidence in favor of learning sophistication.[3]

The current paper conducts classical hypothesis tests of the population rule learning model to demonstrate that sophisticated learning and learning sophistication are empirically significant.

2. **The General Framework and the Rule Learning Model.**

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 JxJ matrix of expected utility payoffs for the row player. For notational convenience, let $p^0 \equiv (1/J, \ldots , 1/J)'$ denote the uniform distribution over A. Further, let p^t denote the empirical frequency of the all players' actions in period t, and 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. Thus, the information available to a representative player at the beginning of period t is $\Omega^t \equiv (G,h^t)$.

The population rule learning model is described fully in Stahl and Haruvy (2002). Briefly, a *behavioral rule* is a mapping from information Ω^t to $\Delta(A)$,

---

[1] Nagel (1995), Ho, *et al* (1998), Duffy and Nagel (1997), Stahl (1996, 1999a,b, 2000, 2001), and Sosino, *et al* (1998).

the set of probability measures on the actions A. Let $\rho \in R$ denote a generic behavioral rule in a space of behavioral rules R.

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 non-negativity restriction on probability measures, the learning dynamics are specified in terms of as the *log-propensity*, $w(\rho,t)$, such that

$$\varphi(\rho,t) \equiv \exp(w(\rho,t)/[\int \exp(w(x,t)dx] . \tag{1}$$

Given a space of behavioral rules R and probabilities φ, the induced probability distribution over actions for period t is $\hat{p}(t) \equiv \int_R \rho(\Omega^t)d\varphi(\rho,t)$.

The third element of the model is the equation of motion:

$$w(\rho,t+1) = \beta_0 w(\rho,t) + \beta_1 \rho(\Omega^t)Up^t, \text{ for } t > 0, \tag{2}$$

where $\beta_0 \in [0,1]$ is a decay parameter and $\beta_1 > 0$ is a scaling parameter.

The space of rules consists of a number of empirically relevant discrete rules that can be combined to span a larger space of behavioral rules. These are the "evidence-based" rules introduced in Stahl (1999a, 2000). The level-1 rule is a logit best reply (LBR) to adaptive expectations of other players, where the latter is given by

$$q^t(\theta) = (1-\theta)q^{t-1}(\theta) + \theta p^{t-1} , \quad \text{for } t \geq 1, \tag{3}$$

and $q^0(\theta) \equiv p^0$.

The level-2 rule is a LBR to the belief that others use the level-1 rule. The Nash rule is a LBR to the belief that each Nash equilibrium action is equally likely. Other rules emerge from convex combinations of these focal beliefs, with weights given by $\nu \equiv (\nu_1, \nu_2, \nu_3)'$. The initial log-propensity function for the evidence-based rules is specified as a normal distribution with the mean $(\bar{\nu}, \bar{\theta})$, and variance σ^2.

In addition to these evidence-based rules, there is a herd rule that follows $q^t(\theta)$ interpreted as: with probability $(1-\theta)$ the player repeats his/her choice propensities from last period, and with probability θ the player mimics the population distribution of choices. The initial propensity to "follow the herd" is denoted by δ. Finally, there is a tremble rule that produces uniformly random choices, with an initial propensity denoted by ε.

Since the data for this paper come from experiments with one run with one game for T periods, followed by a second run with a different game for T periods, we need to specify the initial propensities for the first period of the second run. We use a one-parameter specification:

$$w(\rho, T+1) = (1-\tau)w(\rho, 1) + \tau w(\rho, T^*) , \tag{4}$$

where T* indicates the update after period T of the first run, and τ is the *transference* parameter.

The theoretical model involves 10 parameters: $\xi \equiv (\delta, \varepsilon, \bar{\nu}_1, \bar{\nu}_2, \bar{\nu}_3, \bar{\theta}, \sigma, \beta_0, \beta_1, \tau)$. The rule propensities and law of motion yield population choice probabilities:

$$\hat{p}_j^t(\xi) \equiv \int_R \hat{p}_j(\Omega^t; \rho)\varphi(\rho, t \mid \xi) d\rho , \tag{5}$$

from which the log-likelihood of the data can be calculated.

3. The Experimental Data and Methodology.

The population rule learning model is confronted by data from two different experiments, both using binary lottery games. The first experiment consisted of four sessions, each with two runs of 15 periods and one symmetric 5×5 game in each run. The second experiment consisted of four sessions, each with two runs of 12 periods and one symmetric 3×3 game in each run. For both experiments, in each period each participant was matched with the n-1 other participants and was given the history of choices of those n-1 other players. The binary lotteries were not resolved until all periods of both runs were completed. The games and choice data can be obtained at www.eco.utexas.edu/faculty/Stahl/experimental.

We estimate the parameters of the models using the maximum likelihood approach. To separate out the first period effects from the dynamic effects, we compute the log-likelihood (LL) value for all but the first period of each run, denoted LL(-1). As another measure of fit, we compute the Root Mean Squared Error (RMSE) summed over all periods based on the Euclidean distance between the empirical choice frequences, p^t, and the predicted choice probabilities, $p^e(t)$. We also compute the Pearson Chi-Square (PX^2) measure of goodness-of-fit.

4. Maximum Likelihood Estimates and Hypothesis Tests.

The maximized log-likelihood (LL) function for the entire pooled data set is -4720.09. Table I displays the maximum likelihood parameter estimates. Note that the estimates of three parameters fall on the boundary of the parameter space: $\bar{\nu}_3 = 0$, $\beta_0 = 1$, and $\tau = 1$. We interpret this finding as indicating that a model with just seven "interior" parameters can fit the data as well as the full ten parameter model.

Table II displays four measures of in-sample goodness of fit: (i) the LL of the holdout data given the parameter estimates, (ii) the LL excluding the first period of each run, LL(-1), (iii) the root mean sequated error (RMSE) of the forecast, and (iv) the Pearson Chi-square ($P\chi^2$) goodness-of-fit measure.

Three alternative models are nested within our revised rule learning model. First, by restricting the learning parameter $\beta_1 = 0$, we eliminate rule learning completely, leaving the "Diverse Worldly LBRIAE" model. Second, by further restricting the variance parameter $\sigma^2 = 0$, we eliminate diversity in the population, leaving the "Worldly LBRAE" model. Third, by further eliminating all evidence except level-1 evidence ($\bar{\nu}_2 = \bar{\nu}_3 = 0$), we obtain the "LBRIAE" model.[4] The corresponding in-sample performance of each of these restricted models is reported in Table II.[5] Clearly each of these hypotheses can be rejected at all commonly used acceptance levels.

LL(-1) demonstrates that the improvement in LL is not just a first-period phenomena. While the RMSE measures are not substantially different, they are consistent with the other measures, and the Pearson Chi-square statistic, which is much more sensitive than RMSE, strongly reinforces our findings. Therefore, we conclude that sophisticated learning and learning sopistication are empirically significant.

5. **What Rules are Learned.**

The grid for (ν_1, ν_2 and θ) used to perform the integration required by eq(5) consists of 75 points: 5 points in the θ dimension and 15 points for

---

[2] In Stahl (1999b), "action reinforcement" learning models (replicator dynamics, Roth-Erev reinforcement, Camerer-Ho EWA dynamics, and LBRIAE) were pitted against each other in a horse race. LBRIAE was the clear winner.
[3]Out-of-sample performance is similar (Stahl, 1999b).

(ν_1, ν_2). Given the estimated parameters of the rule learning model, we can compute the propensities, $\varphi(\nu,\theta;t)$, for each grid point over time for each experimental session. The change in $\varphi(\nu,\theta;t)$ is continuous over time and remarkably similar over sessions. From the first to the last period of a session (i) trembles decline dramatically, (ii) the herd rules decline in importance, (iii) the level-1 rules and the rules that combine level-1 and level-2 evidence increase in importance, and (iv) the inertia in adaptive expectations declines (mass shifts to higher θ values).

Computation of the predicted choice frequencies of the handful of rules with non-negligible propensities reveals that these rules are quite distinct period by period. Thus, it is not surprising that the changing φ distribution over these rules produces richer behavioral dynamics than can be generated by the simpler models nested within the rule learning model.

REFERENCES

Anderson, S., J. Goeree, and C. Holt (1997), "Stochastic Game Theory: Adjustment to Equilibrium Under Bounded Rationality," Department of Economics, University of Virginia, mimeo.

Camerer, C, and T. Ho (1997), "EWA Learning in Games: Preliminary Estimates from Weak-Link Games," in R. Hogarth (ed), *Games and Human Behavior: Essays in Honor of Amnon Rapoport*.

________ (1999), "Experience-Weighted Attraction Learning in Normal-Form Games, *Econometrica*, **67**, 827-874.

Cheung, Y-W, and D. Friedman (1997), "Individual Learning in Games: Some Laboratory Results," *Games and Economic Behavior*, **19**, 46-76.

________ (1998), "A Comparison of Learning and Replicator Dynamics Using Experimental Data," *J. of Economic Behavior and Organization*, **35**, 263-280.

Duffy, J., and R. Nagel (1997), "On the Robustness of Behavior in Experimental Beauty contest Games," *Economic Journal*, **107**, 1684-1700.

Erev, I., and A. Roth (1998), "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique Mixed Strategy Equilibria," *American Economic Review*, **88**, 848-881.

Friedman, D., D. Massaro and M. Cohen (1995), "A Comparison of Learning Models," *J. of Mathematical Psychology*, **39**, 164-178.

Ho, T., K. Weigelt, and C. Camerer (1998), "Iterated Dominance and Iterated Best-Response in Experimental 'p-Beauty' Contests", *Amer. Econ. Rev.*, **88**, 947-969.

Kandori, M., Mailath, G., and Rob, R. (1993), "Learning, Mutation, and Long-Run Equilibria in Games," *Econometrica*, **61**, 29-56.

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

________ (1997), "Learning and Decision Costs in Experimental Constant Sum Games," *Games and Economic Behavior*, **19**, 97-132.

Nagel, R. (1995), "Unraveling in Guessing Games: An Experimental Study," *Amer. Econ. Rev.*. **85**, December, 1313-1326.

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

Selten, R. (1990), "Anticipatory Learning in Games," in *Game Equilibrium Models. Vol. I: Evolution and Game Dynamics* (R. Selten, ed), Springer-Verlag.

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

Sonsino, D., I. Erev, S. Gilat, and G. Shabtai (1998), "On the Likelihood of Repeated Zero-sum Betting by Adaptive Human Agents," working paper, Faculty of Industrial Engineering and Management, Technion, Israel.

Stahl, D. (1996), "Boundedly Rational Rule Learning in a Guessing Game," *Games and Economic Behavior*, **16**, 303-330.

Stahl, D. (1999a), "Evidence Based Rules and Learning in Symmetric Normal-Form Games, *Int'l J. of Game Theory*, **28**, 111-130.

Stahl, D. (1999b), "A Horse Race Among Action Reinforcement Learning Models," March, mimeo.

Stahl, D. (2000), "Rule Learning in Symmetric Normal-Form Games: Theory and Evidence, *Games and Economic Behavior*, **32**, 105-138.

Stahl, D. (2001), "Population Rule Learning in Symmetric Normal-Form Games: Theory and Evidence," *J. of Econ. Org. and Behavior*, **45**, 19-35.

Stahl, D. and E. Haruvy (2002), "Aspiration-based and Reciprocity-based Rules in Learning Dynamics for Symmetric Normal-Form Games," *J. Math. Psych.* forthcoming.

Table I. Maximum Likelihood Parameter Estimates

Parameter	Population Rule Learning	Diverse Worldly LBRIAE	Worldly LBRIAE	LBRIAE
δ	0.533	0.516	0.532	0.573
$\underline{\varepsilon}$	0.090	0.033	0.059	0.053
$\underline{\nu}_1$	0.797	0.968	0.358	0.400
$\underline{\nu}_2$	0.070	0.092	0.022	
$\underline{\nu}_3$	0.000	0.000	0.000	
θ	0.647	0.641	0.816	0.749
σ	0.767	0.757		
β_0	1.000			
β_1	0.008			
τ	1.000			
LL	-4720.09	-4730.79	-4756.04	-4772.52
No. of interior parmeters	7	6	5	4

Table II. In-Sample Performance Measures

Model	LL	LL(-1)	RMSE	PX^2
Pop Rule Learning	-4720.09	-4292.34	0.0790	834.6
Div Worldly LBRIAE	-4730.79	-4305.20	0.0800	866.5
Worldly LBRIAE	-4756.04	-4329.20	0.0829	937.0
LBRIAE	-4772.52	-4338.17	0.0859	993.7

(for reviewer's convenience - not for publication)

Appendix A: Out-of-Sample Performance Measures

Model	LL	LL(-1)	RMSE	$P\chi^2$
Pop Rule Learning	-4754.17	-4325.98	0.0834	909.3
Div Worldly LBRIAE	-4760.96	-4335.62	0.0833	936.3
Worldly LBRIAE	-4793.51	-4368.02	0.0872	1048
LBRIAE	-4794.42	-4359.11	0.0878	1068

(for reviewer's convenience - not for publication)

APPENDIX B. Table II from Stahl (1999b)

Model	Deviations	np	LL	LL(-1)	RMSE	$P\chi^2$
Mental Replicator						
ratio form:	both	3	-5386.62	-4918.46	.161	2290
logit form:	both	3	-4992.73	-4563.06	.118	1437
Roth-Erev						
uniform prior:	mutations	7	-4916.93	-4402.76	.107	1305
insuff. reason:	mutations	7	-4828.06	-4405.53	.0941	1129
LBR	trembles*	3	-4805.13	-4369.35	.0882	1033
EWA						
power form:	trembles*	5	-4796.04	-4366.37	.0904	998
logit form:	trembles*	6	-4783.64	-4353.29	.0892	968
logit form:	trembles*	4	-4784.07	-4353.28	.0894	966
LBR*	trembles*	4	-4772.52	-4338.17	.0859	994
Rule-Learning		6	-4737.40	-4314.32	.0814	971

*These models implicitly contain a form of mutations.
"both" = mutations and trembles

(for reviewer's convenience - not for publication)

APPENDIX C: Payoff Matrices of Experiment Sessions

Session	Run 1					Run 2				
	68	10	76	33	75	19	43	96	85	85
	73	4	59	0	8	28	62	88	74	24
6/27/95	3	92	16	15	99	67	21	38	48	38
	86	54	25	41	6	40	58	0	15	92
	72	98	92	8	52	16	15	86	99	79
	2	31	0	99	6	22	79	35	56	75
	6	10	97	40	24	22	38	78	55	99
6/29/95	98	96	38	48	19	27	58	1	11	0
	42	40	80	51	48	70	1	34	59	37
	97	46	5	68	49	56	84	60	23	2
	22	79	35	56	75	2	31	0	99	6
	22	38	78	55	99	6	10	97	40	24
8/10/95	27	58	1	11	0	98	96	38	48	19
	70	1	34	59	37	42	40	80	51	48
	56	84	60	23	2	97	46	5	68	49
	19	43	96	85	85	68	10	76	33	75
	28	62	88	74	24	73	4	59	0	8
8/15/95	67	21	38	48	38	3	92	16	15	99
	40	58	0	15	92	86	54	25	41	6
	16	15	86	99	79	72	98	92	8	52

Session	Run 1			Run 2		
2/17/98	20	0	60	30	50	100
	0	60	0	40	45	10
	10	25	25	35	60	0
2/19/98	80	60	50	30	100	22
	60	70	90	35	0	45
	0	0	100	51	50	20
4/07/98	50	0	0	70	60	90
	70	35	35	60	80	50
	0	25	55	40	20	100
4/09/98	60	60	30	68	4	49
	30	70	20	86	41	4
	70	25	35	72	25	39