# ON THE POTENTIAL USES AND CURRENT LIMITATIONS OF DATA DRIVEN LEARNING MODELS[1]

Ido Erev[a], and Ernan Haruvy[b]

[a]Columbia Business School and Technion

[b]Harvard Business School and Technion

January 2001

**Abstract**

The experimental study of human adjustment to economic incentives has been deadlocked for quite some time by apparently contradictory conclusions as to which is a better theory of learning. This article attempts to shed some light on this impasse by pointing out that different learning models often have different objectives that imply different model comparison criteria. The different criteria are expected to lead to the same conclusions if the models are perfectly specified, but might lead to different conclusions when they are used to compare approximations. We discuss the potential usefulness of learning models in light of the observation that they are likely to be misspecified, and outline the type of applications appropriate for each approach.

*Keywords*: Learning; Model Comparison; Evaluation Criteria

*JEL classification*: D83; C91

[1] We are especially indebted to Brit Grosskopf, Dale Stahl and Al Roth for helpful comments and corrections. All remaining errors and omissions are of course our sole responsibility.

## 1. Introduction

A comparison of recent studies that examine human adjustment to economic incentives reveals a disconcerting picture. Different studies appear to contradict each other. Some studies (e.g., Roth and Erev, 1995; Erev and Roth 1998; Sarin and Vahid, 2000) suggest that learning is best approximated by simple reinforcement models. Another line of research (Camerer and Ho, 1997, 1998, 1999a, 1999b) argues that choice reinforcement models can be rejected in favor of more general experience weighted attraction (EWA) models. Yet a third study (Stahl, 1999a) shows that both reinforcement and EWA models are outperformed by a simple logit best reply with inertia and adaptive expectations along the lines of the stochastic fictitious play model studied by Fudenberg and Levine (1998)[2]. Recently, Feltovich (2000) showed that the rankings of simplified models were dependent on the specific games, the assumed parameters, and the specific measures of goodness of fit used. In light of these apparent contradictions and lack of convergence on a single model, the distinct descriptive learning studies can be criticized for being closer to descriptions of religions than to scientific research. An outsider to learning could falsely conclude that the study of human adaptation has yet to discover robust regularities.

The main goal of the current paper is to try to clarify the reasons for these apparent inconsistencies. We begin, in section 2, with the observation that although all studies of the effect of experience on economic behavior share the same long-term goal— namely, understanding economic behavior-- the routes taken to arrive at that goal are quite distinct. Camerer and Ho (1999b), using traditional one-period-ahead goodness of fit techniques, allow for the possibility of different parameters in each game. Stahl (1999a) and Cheung and Friedman (1998), on the other hand, relying as well on one-period-ahead techniques, insist on a single set of parameters over games. Roth and Erev (1995), Erev and Roth (1998), Sarin and Vahid (2000), and Goeree and Holt (1999) likewise assert a single set of parameters for a class of games, yet their focus is on simulation (T-period-ahead) measures. Under the assumption that the different models are well specified, the different criteria are not expected to result in different rankings over models. However, human behavior is affected by unobserved interactions between billions of neurons and many environmental factors, and it is unlikely that a model of a few parameters can perfectly capture behavior, nor would scientists

[2] Fudenberg and Levine (1998) present a comprehensive analysis of the long-term properties of several models including a variant of the reinforcement model studied by Erev and Roth (1998). The current paper focuses on the "intermediate" and short-term predictions of these models.

generally attempt to perfectly specify any phenomenon[3]. This observation implies that the different conclusions reached by different researchers are possibly a result of the different criteria used to compare models.

Section 3 presents a numerical example in support of the above suggestion. Specifically, we generate data from a pre-specified model with pre-specified parameters. We then estimate parameters for a "misspecified" model which pools parameters over individuals. The estimation is performed using two different criteria from the literature. The resulting parameter estimates under each approach are strikingly different; yet each is meaningful and useful for a different purpose.

Section 4 explores our assertion experimentally. It presents a simple data set that can support three apparently contradicting conclusions. A focus on Camerer and Ho's (1999a,b) criteria appears to suggest that the experience-weighted attraction model proposed in their works has the best fit in terms of game-by-game one-period-ahead likelihood. A focus on simulation-based statistics shows a model proposed by Erev et al. (1999) leading. A pooled-game one-period-ahead maximum likelihood analysis shows that yet a third model, based on Stahl (1999a) best fits the data.

Section 5 concludes with a discussion of the potential usefulness of learning models in light of the observation that they are most likely misspecified. We argue that this observation does not imply that the various models cannot be useful, but rather that the criterion used to select a model should depend on the intended use of that model.

## 2. The distinct routes of data driven learning research

Most studies of human adjustment to economic incentives were designed to contribute to the same long-term goal: the development of a general descriptive theory of economic behavior. Yet, different researchers take very different routes in the quest for the holy grail of learning research. For the current discussion, it is important to distinguish between alternative data driven routes. We classify routes along two dimensions: (1) the generalizability of parameters across games, and (2) the focus on one-period-ahead versus T-period-ahead predictions. It should be emphasized, however, that there is a third dimension

[3] Occam's razor is often cited by advocates of simple models to capture complex phenomena: "We could still imagine that there is a set of laws that determines events completely for some supernatural being, who could observe the present state of the universe without disturbing it. However, such models of the universe are not of much interest to us mortals. It seems better to employ the principle known as Occam's razor and cut out all the features of the theory which cannot be observed" (Hawking, 1988).

that will receive less attention in the current paper, yet is no less important than the other two. That is, the generalizability of parameters over individuals. Just as behavior differs across different games, so it does across different players. Studies such as Daniel, Seale, and Rapoport (1998), Rapoport, Daniel and Seale (1998), Cheung and Friedman (1997), Stahl (1996), and Camerer, Ho and Wang (2000) concluded that individuals are sufficiently different that pooling them together implies a grave misspecification.

The fact that the many of the above authors in subsequent papers choose to nonetheless pool over individuals puzzles many readers, as eloquently expressed by Nyarko and Schotter (2000): "It is ironic that while [Erev-Roth, 1998, and Camerer-Ho, 1999] are formulated as models of individual learning, when they are tested and compared, the authors too often aggregate the data, either across time or across individuals, and make their comparisons on the aggregate level."

Despite our ability to occasionally capture heterogeneity with clever models[4], such models can easily get out of hand and lose robustness, particularly when the complexity of learning behavior is added. Hence, despite convincing evidence in Cheung and Friedman (1997), Stahl (1996), and Camerer and Ho (2000) against aggregation over individuals, studies such as Cheung and Friedman (1998), Stahl (2000a, 2000b), and Camerer and Ho (1999b) succumb to the need for parsimony and pool parameters over individuals. This rare concensus in regard to individual parameters prompts us to investigate the other two dimensions with greater care.

### 2.1. One-period-ahead without enforcing parameter generalizability over games

The purest approach to analyze data calls for using all the observations and for testing all the assumptions that can be tested. In the context of learning research this approach implies a focus on one period ahead (predicting each observation based on all the data that could affect it), and statistical tests of the significance of all the relevant variables and interactions. The one-period-ahead prediction minimizes a likelihood function[5] defined by

[4] One approach suggested by Stahl and Wilson (1994, 1995), is to divide players into sub-populations of similar characteristics. Even then, Haruvy, Stahl, and Wilson (2000) suggest that heterogeneity within each sub-population should be modeled with extreme caution.

[5] We present here the representative agent approach. Other approaches involve heterogeneous population models (Stahl, 2000) and individual player estimation (Cheung and Friedman, 1997; Stahl, 1999b).

$$L = \prod_{i=1}^{N} \prod_{t=1}^{T} P_{it}(x_{it} \mid x_1, \ldots, x_{t-1}, \beta), \quad (1)$$

where i indexes players, t indexes time periods, $x_t$ is the choice of player i at time t, $\beta$ is the vector of the model's parameters, and $P_{it}$ is the one-period-ahead probabilistic prediction for player i's choice based the model's parameters and on all the choices, by all the players, up to and including period t-1. $P_{i1}$ is either assumed or estimated and its treatment is beyond the scope of this paper.

In reality there are more assumptions than can be reasonably tested and the pure approach cannot be utilized without the imposition of some constraints. All that data driven researchers can do is try to reduce the number of constraints in the analysis. Camerer and Ho (1997, 1998, 1999a, 1999b) provide some of the most prominent examples of data driven analyses of this type. They propose a general model of learning, *experience weighted attraction* (EWA), and use one-period-ahead econometric analysis, along the lines of eq(1), to estimate its parameters and select the significant explanatory variables. Their results clearly show that three of the constraints imposed by other researchers can be rejected: (1) Decision-makers are affected by forgone payoffs (in violation of pure reinforcement learning models). (2) The effect of forgone payoffs tends to be weaker than the effect of realized payoffs (in violation of pure belief-learning models). (3) The assumption of general parameters across games can be rejected.

### 2.2. One-period-ahead with parameter generalizability over games

Despite the general statistical finding that games cannot be pooled (e.g., Stahl, 1996), it is nevertheless not unreasonable to analyze data sets under the constraint of general parameters over games. This constraint is often imposed (e.g., Stahl, 1999a; Cheung and Friedman, 1998) to facilitate ex-ante predictive ability. As noted by Feltovich (2000), regarding Camerer and Ho's (1999b) approach: "… when they find that a particular combination of parameters best fits a set of experimental data, questions come to mind concerning how sensitive these best parameters are to small changes in the game, whether one could predict which parameter values are appropriate for which games and so on." If games have different parameters, one's ability to predict ex-ante dynamic play in a new setting is limited. Generalizability allows estimating one set of parameters on a large number of games and using this set of parameters for ex ante predictions in similar games.

The research conducted by Stahl (1999a) arrived at the *logit best reply with inertia and adaptive expectations (LBRIAE)*. Stahl (1999a) compared the leading learning models and found that this LBRIAE model best fits the data from eight games (under the constraints of a single set of parameters in all eight games). The LBRIAE model is a sophisticated variant of the stochastic fictitious play model studied by Fudenberg and Levine (1998). In LBRIAE, a player's probability distribution in period t is a weighted average of the logit best-response mapping from a gradually updated prior with a precision parameter, an imitation of the population's play in period t-1, and a uniform tremble. As players in our treatments are not matched against others and hence receive no information about the aggregate empirical frequency, we replace the imitation component of Stahl (1999a) with the player's own historical frequency of choice. Though, due to lack of creativity on our part, we select to keep the name LBRIAE for this modified version, we must caution that replacing 'population imitation' with 'individual inertia' may have drastic conseuqneces. For the correct implementation of LBRIAE, the reader is encouraged to refer to Stahl (1999a).

### 2.3. T-periods-ahead with parameter generalizability over games

Some researchers in the field of learning (e.g., Erev and Roth, 1998; Van Huyck et al., 1997; Goeree and Holt, 1999; Sarin and Vahid, 2000) choose to compare model simulations of the entire paths to observed paths as opposed to period-to-period predictions. This approach to study learning appears econometrically deficient. Under this approach the researchers use computer simulations in an attempt to find the model that best predicts the observed learning curves (T-periods ahead) in a set of games. This approach appears inefficient because the individual history of each of the different subjects is not considered in the analysis, despite the fact the entire purpose of the model is to capture history-dependent behavior.

Given that models are misspecified, however, this approach is not at all unreasonable. In a misspecified model, since the model is a rough approximation of behavior, minimizing predictive error one-period-ahead does not necessarily guarantee a minimization of a T-period-ahead prediction. Recall the notation of section 2.1. Using that notation, the appropriate likelihood function corresponding to the t-period ahead prediction of choices $x_t$, unconditional of $x_1$ through $x_{t-1}$ would be

$$L_t = \boxed{\times} \int \ldots \int f_{it}(x_{it} \mid x_1, \ldots, x_{t-1}, \beta)\ df_1 \ldots df_{t-1} \qquad (2)$$

where $f_t$ is the density of choices $x_t$ at time t. Hence, if all periods t are to be given equal weight as in eq(1), the appropriate unconditional likelihood would be

$$L = \boxed{\times} L_t \qquad (3)$$

One problem in this approach is arriving at the appropriate density representation, $f_t$. The second problem is that the number of integrations required here is computationally infeasible. Haruvy (1999) suggests using simulations, combined with kernel density estimation, to arrive at the unconditional likelihood. A large number of T-period-ahead simulations would give us the predicted T-period-ahead density of frequencies necessary to approximate the unconditional prediction each period.

With the T-periods-ahead simulation, if the final outcome distribution is unimodal, then the simulation mean is an informative statistic, and is well captured by the MSD approach. One can think of ordinary least squares as identifying the mean with the sum of squared errors indicating the variance around the mean. In this case, the MSD criterion will always prefer the model that is closest in the mean but has minimal variance from the mean. The only "error," by this approach, is in the variance, and the MSD measure itself provides the "correction."[6]

Erev, Bereby-Meyer, and Roth (1999) have pursued the T-period-ahead MSD approach with their two-parameter reinforcement learning model, hereafter referred to as REL (for REinforcement Learning). In 80 different tasks, REL has performed satisfactorily by the MSD criterion, with the same two parameters. Hence, in the remainder of the paper, we will analyze two versions of REL-- one with one-period-ahead maximum likelihood parameters and the other with the fixed parameters recommended by Erev, Bereby-Meyer, and Roth (1999).

[6] However, if the final outcome is multimodal (as in the separatrix-crossing games of Haruvy and Stahl, 1999), then capturing the mean could be potentially inadequate, as the true distribution would not be well represented by the unimodal MSD approach. To capture the observed distribution in these cases the MSD criterion has to be replaced with a measure of the difference between the predicted density function over outcomes with the empirical density function over outcomes, but that of course requires many more empirical observations than are

### 2.4. The relationship between the different routes

Econometric theory implies that under certain conditions the differences between the three routes described above do not require distinct research methods. Most importantly, under the assumption of a well-specified research model, all three approaches are expected to be appropriate for parameter estimation, significance testing, model evaluation, and model selection. On the other hand, if the models are not perfectly specified, the different routes of learning research may lead to different conclusions. The experiment reported below was designed to evaluate whether these possible differences could be a sufficient reason for the inconsistencies described above.

## 3. A Numerical Example

Given the misspecification inherent in models which pool parameters over players or games, the different criteria and interpretation assigned to parameters take on new importance. To demonstrate the need for caution in choosing evaluation criteria and parameter interpretation, Haruvy and Erev (2000) propose the following binary choice task: Assume that the probability of choice A by player i at trial t is

$$P_i(t) = \alpha\, F_i(t-1) + (1-\alpha)(0.3 + \varepsilon_i), \qquad (4)$$

where $P_i(1) = F_i(1) = 0.5$, and for $t > 1$, $F_i(t) = [\, F_i(t-1)\cdot(t-1) + \chi_i(t)\, ] / t$, $\chi_i(t)$ takes on the value of 1 if A is chosen by player i at time t and 0 otherwise, and $\varepsilon_i$ is uniformly distributed between –0.3 and 0.3. In other words, a player i adjusts sluggishly towards the propensity of his type, where types are uniformly drawn between 0 and 0.6. Suppose we know that the average player will eventually choose A with a probability of 0.3, but we don't know the value of the parameter $\alpha$. One natural interpretation is that $\alpha$ is *the speed of adjustment*. When $\alpha$ equals 1, the player never adjusts, and when $\alpha$ equals 0, the player adjusts immediately to the final propensity corresponding to his type. However, though we are aware of heterogeneity, we select to conform to the "rare consensus" mentioned in section 2 with regard to pooling parameters over individuals. Hence, we will not estimate the epsilon parameter for each player. Since we know that epsilon is on average 0, we eliminate epsilon

---

available in most experiments. The current discussion focuses on situations with a dominant strategy (or a unique equilibrium) and hence the multimodality problem is not expected to play a key role.

from the estimated model. We then estimate alpha. To examine whether $\alpha$ retains its original interpretation we first generate an artificial data set using simulations with a true value of $\alpha$ = 0.3. The left-hand side of Figure 1 presents the proportion of A choices of simulated subjects in this setting. The parameter $\alpha$ is then estimated using the two common estimation methods in learning research. The first estimation method, T-*period-ahead with generalizability* (TPG), is a T-period-ahead simulation approach with the minimum mean squared distance criterion. By that approach, one simulates hundreds of players under different sets of parameter values, aggregates the results and compares aggregates. The second method, *one-period-ahead with generalizability* (OPG), is a one-period-ahead likelihood approach. Under that approach, each period, based on the player-specific past, the model produces a player-specific prediction for the following period, which is then compared to the actual player choice, thereby constructing a likelihood function. The estimated values are $\alpha$ = 0.19 in the TPG method and $\alpha$ = 1 in the OPG method. The right-hand side of Figure 1 presents the aggregated predictions of the model (derived using computer simulations) using these two estimates.

The results clearly show that the OPG estimate clearly fails to reproduce the adjustment process though both estimates are incorrect. Nevertheless, it is important to see that both estimates are informative. Specifically, the OPG result ($\alpha$ = 1) implies that given the heterogeneity in the population, an individual player's past frequency of choice is the best predictor of his next period choice. This finding is informative, though $\alpha$ can no longer be thought of as the speed of adjustment parameter.

Looking at the longer-horizon TPG approach, we see that $\alpha$ is estimated at 0.19. Since the TPG approach in essence aggregates over individuals (a simulation approach ignores individual histories), $\alpha$ could conceivably be interpreted as the speed of adjustment parameter for the representative player. Yet, the probability that this estimate is correct for a randomly picked individual would be nearly nil.

Although aggregation over players, as well as the method of estimation chosen, were shown to be consequential for the interpretation of parameters, the lack of aggregation may be no less perilous for interpretation. Individual player estimates may carry little information for prediction about the population. Similarly, individual game estimates may carry little predictive power and reek of overfitting.

## 4. An Experimental Demonstration

We selected to study a set of simple choice tasks for which the different models have different predictions. Though the tasks described here may seem rather simple, if the models are not well specified for simple tasks they are not likely to fare better in complex tasks.

The first task we considered was a choice between two sure gains: 10 tokens versus 11 tokens (100 tokens = 1 shekel). Since the different models address losses differently we also studied a loss version of this task: a choice between a sure loss of –10 and a sure loss of –11. To allow evaluation of the effect of payoff variance (another important difference between the models), two noisy variants of the basic problems were added. Noise was introduced by replacing the outcome of 11 (-11) with a gamble that pays 1 and 21 (-1 and –21) with equal probabilities. The four conditions are summarized on the left hand-side of Figure 2. We name the conditions by the payoff to the button that is different between all treatments. Hence we have condition (11), condition (-11), condition (1, 21), and condition (-1, -21).

### 4.1 Method

A total of 40 subjects participated in the study. They were recruited by ads that were posted around the Technion campus and promised a substantial amount of money for a participation in a short decision making experiment. Subjects were randomly assigned to one of four experimental conditions, were seated at separate terminals and presented with two buttons on the computer screen. Subjects were told that the experiment consisted of 200 trials and their task was to select between the two buttons. A number would appear on each button after they made their selection, and the number on the selected button would be added to their payoff.

### 4.2 Results

Figure 2 summarizes the main experimental results. It presents the proportions of choices of the button with the 10 (or –10) outcome in the four experimental conditions (the data are grouped into four blocks of 100 trials). Not surprisingly, the results reveal quick learning to maximize earnings in the certain outcome conditions, and substantially slower learning in the probabilistic outcome conditions.

Table 1 summarizes the results of a three-model comparison. Notice that in addition to the three models with maximized values, we calculate likelihood for REL with non-maximized parameters, as Erev, Bereby-Meyer and Roth (1999) found a set of parameters

that provided robust predictions in 80 decision tasks. We are interested in comparing their parameters' predictive power here as well.

The o*ne-period-ahead with no generalizability* (OPNG) approach examines which of the three models presented above minimizes the total likelihood treatment-by-treatment. EWA has the highest (best) log-likelihood in all but condition (-11), with log-likelihoods of -362.559, -397.295, -1205.31, -1237.91 for conditions (11), (-11), (1, 21), and (-1, -21), respectively. It would seem that EWA is a horse-race winner relative to LBRIAE (-386.348, -374.136, -1225.13, -1245.66) and REL (-378.918, -349.495, -1348.31, -1309.15).

The o*ne-period-ahead with generalizability* (OPG) approach examines which of the three models presented above minimizes the total likelihood for the pooled treatments. LBRIAE dominates the other two with a pooled log-likelihood of -3098.70, as compared to EWA (-4706.87) and REL (-3398.32).

The T-*period-ahead with generalizability* (TPG) approach examines which of the three models presented above minimizes the total MSD of simulated paths from actual paths for a single set of parameters. We derive that set of parameters from the pooled–treatments maximum likelihood (one-period ahead ML due to both the computational complexity of the alternative and the relatively small number of observations). For REL, we also derive the MSD statistic for the fixed REL parameters from Erev et al (1999). Both variants of REL dominate the other two models, EWA and LBRIAE, for three out of four treatments. In condition (–1, -21), only the ML variant of REL dominates both EWA and LBRIAE. The results of these simulations are presented on the right hand side of Figure 2.

The current results imply that the three models studied here are not well specified, and that this mispecification can have a large effect to the extent that different routes lead to very different conclusions.

## 5. Conclusion: The usefulness of learning models

We have demonstrated with one set of simple individual decision tasks that different goodness of fit criteria for learning models may imply different horse-race winners. In our experiment, these winners, not surprisingly, are consistent with the seemingly different conclusions of the various works applying each approach. The interpretation we assign this finding is that the various models are not well specified. It seems that in the current setting, econometric analysis can be very sensitive to this misspecification.

Our findings are both good news and bad news. The bad news is that the learning models we considered are most likely misspecified, and that there is not likely to be a "correct" econometric criterion to compare them. The good news is that results of previous research on learning no longer seem contradictory. The fact that the current analysis replicates the results of previous applications of each of the three approaches suggests that there may not be real inconsistencies between the results of previous studies. Rather, this research discovered and quantified three related lines of behavioral regularities. Thus, the models supported in various lines of research are complementary rather than contradictory. However, one needs to be careful of when to apply each approach. For example, in problems of equilibrium selection, period-to-period predictions are not particularly meaningful as one is interested in ex-ante predicting a final outcome. Hence, studies of learning models as predictors of equilibrium, such as Van Huyck et al. (1997), Goeree and Holt (1999) and Haruvy and Stahl (2000) pursue a simulation-based analysis. At the other end of the spectrum, Nyarko and Schotter (2000) pursue one-period-ahead analysis in conjunction with belief elicitation to demonstrate that humans best respond to beliefs and further, that leading models of belief formation in adaptive dynamics are mispecified (in line with this paper's premise). Such an investigation would not be possible at the aggregate path level.

Other examples of the useful complementarity of the two approaches can be seen in various real-life problems. Cosider the decision problem facing the educated casino operator: Casino operators are likely to be interested in questions of the following nature: (1) What will be the long-term effect on casino profits of a particular change in the payoff distributions in the slot machines? (2) What is the effect of the Poker payoffs at time t on the strategy of a particular Poker player at time t+1? (3) How can information concerning learning in Blackjack be used to predict adjustment in a Poker game?

The current results imply that learning research can help answer all three questions, but that three different approaches to modeling may be needed. The TPG approach provides a nontrivial answer to the first question. For example, Haruvy, Erev and Sonsino (2000) show that REL implies that the addition of medium prices is expected to increase gambling. This prediction was validated in an experimental study.

The OPNG modeling approach may run into some difficulty with the first question (as a change in the payoff distribution is expected to alter the model's parameter in an unexpected direction), but it is likely to best answer the second question. For example, the Casino

operators can estimate the EWA parameters for the player of interest and use the model to predict behavior in trial t+1.[7]

The third question is best answered with the OPG approach used by Stahl (1999a). This approach allows for using parameters, estimated in Blackjack, to predict behavior in Poker.

For a more serious example consider the problem of rule enforcement. As in the Casino example, the TPG approach can be used to predict the long term effect of changes in the incentive structure. For example, it has been used to predict the conditions under which enforcement campaigns are expected to succeed (Shany & Erev, 2000), and the value of bad lotteries (like a flash of a red light camera) as punishments (Perry, Haruvy & Erev, 2000). The one-period-ahead approaches are needed to predict immediate responses to punishments.

It should be emphasized that it is not necessarily the models but rather the approaches-- as defined by the goodness of fit and parameter selection criteria-- that are expected to have different strengths in different problems. Though in the experiment at hand the individual models (i.e., EWA, LBRIAE, and REL) have each been shown to be the "horse-race winner" under a different criterion, they may nonetheless provide meaningful and useful predictions to each of the questions above. Using TPG parameters from Bereby-Meyer and Erev (1998) for a variety of models, including belief-based models, and EWA, Haruvy, Erev, and Sonsino (2000) showed that all learning models investigated arrived at the same qualitative predictions in the Casino problem described above.

Moreover, the reader should be cautioned not to form strong conclusions on the salience of the models based on the one simple experiment we presented. The simple tasks were designed to make a point about the limitation and strengths of model comparison and evaluation criteria and not to promote one model over another. For example, despite the flat TPG path of the modified LBRIAE in our experiment, the unmodified LBRIAE in symmetric normal-form games has been shown to be quite sucessful in TPG analysis (Stahl, 1999a) and in TPG equilibrium selection (Haruvy and Stahl, 2000).

Furthermore, the fact that the models studied here were found to be horse race winners dose not imply that these models are the best in their class. Indeed, we hope that the current demonstration of the significance of the differences between the classes will facilitate the

[7]When t is small, the assumption of robust parameters over individuals can be useful. When t is large, the OPNG method can be "improved" by estimating separate parameters for each individual .

design of more efficient model comparison studies and the development of better models for each class.

A second methodological implication of the current demonstration pertains to the relative value of new experimental results. Under the suggestion that learning models are at best useful approximations, the demonstration of a specific violation of a specific model provides only limited information. The best models are the ones that provide the best approximation in the relevant class of tasks. Thus, post hoc models that are proposed to overcome a specific violation of existing models should also be able to capture data previously explained by the models they replace.

In summary, models need not be perfectly specified to be useful. They merely have to capture enough robust regularities to be suitable for a specific purpose. In cases where the purpose is to predict a game path based on previous choices in that same game, the game-specific approach seems preferable for finding suitable models. In cases where short-run predictions need to be made ex-ante, the one-period-ahead generalizability approach is recommended. In cases where long-run predictions are needed ex-ante, a generalizable simulation-based approach is called for.

## References


Bereby-Meyer, Y. and I. Erev (1998), "On Learning To Become a Successful Loser: A Comparison of Alternative Abstractions of Learning Processes in the Loss Domain," *Journal of Mathematical Psychology* **42**, 266-286.

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

Cheung, Y-W, and D. Friedman (1998), "Comparison of Learning and Replicator Dynamics Using Experimental Data," Journal of *Economic Behavior and Organization*, 35, 263-280.

Camerer, C. and T. Ho (1997), "EWA Learning in Games: Preliminary Estimates from Weak-Link Games," in D. Budescu, I. Erev, and R. Zwick (eds), Games and Human Behavior: Essays in Honor of Amnon Rapoport.

Camerer, C. and T. Ho (1998), "EWA Learning in Coordination Games: Probability Rules, Heterogeneity, and Time-variation," *Journal of Mathematical Psychology*, 42:2, 305-326

Camerer, C. and T. Ho (1999a), "Experience-weighted Attraction Learning in Games: Estimates from Weak-link Games," Ch. 3, pp. 31-53 in Games and Human Behavior, David V. Budescu, Ido Erev, and Rami Zwick,e(ds.), Lawrence Erlbaum Associations, Inc.

Camerer, C. and T. Ho (1999b), "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica*, 67, 827-874.

Camerer, C., T. Ho and X. Wang (2000), "Individual Differences in the EWA Learning with Partial Payoff Information," Working paper.

Daniel, T. E., Seale, D. A. & Rapoport, A. (1998), "Strategic play and adaptive learning in sealed bid bargaining mechanism," *Journal of Mathematical Psychology*, 42, 133-166.

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

Erev, I., Y. Bereby-Meyer, and A. Roth (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.

Feltovich, N. (2000) "Reinforcement-Based vs. Beliefs-Based Learning in Experimental Asymmetric-Information Games," *Econometrica* 68, 605-641.

Fudenberg, D. and D. K. Levine (1998) *The Theory of Learning in Games*, MIT Press.

Goeree, Y. and C. Holt (1999), "An Experimental Study of Costly Coordination," University of Virginia working paper.

Haruvy, E. (1999), "Initial Conditions, Adaptive Dynamics, And Final Outcomes," Doctoral Dissertation, University of Texas.

Haruvy, E., I. Erev and D. Sonsino (2000), "The Medium Prizes Paradox: Evidence From a Simulated Casino," forthcoming in *Risk and Uncertainty*.

Haruvy, E. and I. Erev (2000), "On the Application and Interpretation of Learning Models," Technion working paper.

Haruvy, E. and Stahl, D. O. (1999), "Initial Conditions, Adaptive Dynamics, And Final Outcomes," University of Texas working paper.

Haruvy, E. and Stahl, D. O. (2000), "Deductive versus Inductive Equilibrium Selection: Experimental Results," University of Texas working paper.

Hawking, S. (1988), A Brief History of Time, Bantam Books, New York.

Ho, T., C. Camerer, and K. Weigelt (1998), "Iterated Dominance and iterated Best-Response in Experimental 'P-Beauty Contests,'" *The American Economic Review*, 88:4, 947-969.

Nyarko, Y. and Schotter, A. (2000), "An Experimental Study of Belief Learning Using Elicited Beliefs," NYU working paper.

Perry, O., E. Haruvy, and I. Erev (2000), "Frequent Delayed Probabilistic Punishment in Law Enforcement," forthcoming in *Economics of Governance*.

Rapoport, A. Daniel, T. E. and D. A. Seale (1998). "Reinforcement-based adaptive learning in asymmetric two-person bargaining with incomplete information," *Journal of Experimental Economics,* **1,** 221-253.

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.

Sarin, R. and F. Vahid (2001) "Predicting How People Play Games: A Simple Dynamic Model of Choice," *Games and Economic Behavior*, 34 (2001), 104-122.

Shany, D. and I. Erev (2000), "On the Rule Enforcement Paradox and the Solution Implied by Models of Reinforcement Learning," Technion working paper.

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

Stahl, D. O. (1999a), "A Horse Race Among Action Reinforcement Learning Models," University of Texas working paper.

Stahl, D. O (1999b) "Evidence Based Rule Learning in Symmetric Normal-Form Games," *International Journal of Game Theory*, 28, 111-130, 1999.

Stahl (2000) Population Rule Learning in Symmetric Normal-Form Games: Theory and Evidence," *J. of Economic Behavior and Organization*, forthcoming

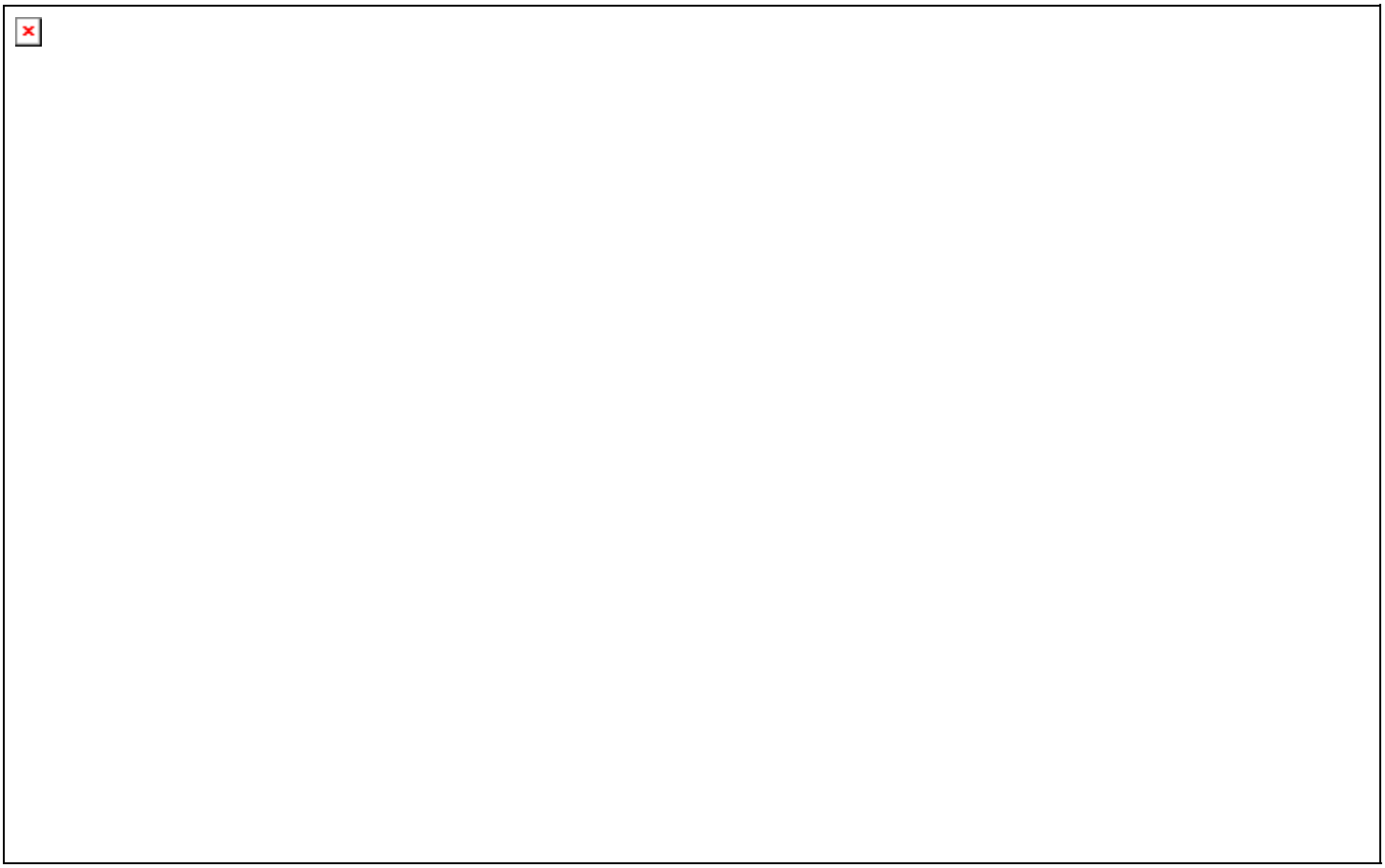
Van Huyck, J., J. Cook, and R. Battalio (1997), "Adaptive behavior and coordination failure," *Journal of Economic Behavior & Organization*, **32** (4). 483-503.

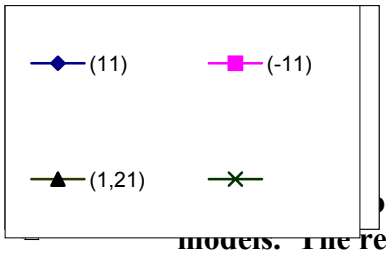
**TABLE 1: Summary of the three criteria:** The last column has REL with parameters from Erev et al (1999). All other likelihood results are with maximum likelihood parameters corresponding to the treatment under consideration.

| Criterion | Condition | EWA | LBRIAE | REL | REL (fixed) |
|---|---|---|---|---|---|
| LL with game specific parameters | 11 | -362.56 | -386.35 | -378.92 | -498.49 |
| | -11 | -397.30 | -374.14 | -349.50 | -477.54 |
| | 1,21 | -1,205.31 | -1,225.13 | -1,348.31 | -1,376.17 |
| | -1,-21 | -1,237.91 | -1,245.66 | -1,309.15 | -1,353.26 |
| | Mean | **-800.77** | -807.82 | -846.47 | -926.365 |
| LL with general parameters | 11 | -872.70 | -386.35 | -379.41 | -498.49 |
| | -11 | -868.82 | -374.14 | -351.61 | -477.54 |
| | 1,21 | -1444.22 | -1207.57 | -1355.68 | -1376.17 |
| | -1,-21 | -1521.13 | -1130.65 | -1311.63 | -1353.26 |
| | Pooled | -4,706.87 | **-3,098.70** | -3,398.32 | -3,705.46 |
| MSD (for 5-blocks of 40 periods) with one set of parameters | 11 | 0.057 | 0.081 | 0.014 | 0.011 |
| | -11 | 0.088 | 0.209 | 0.063 | 0.018 |
| | 1,21 | 0.061 | 0.062 | 0.055 | 0.051 |
| | -1,-21 | 0.086 | 0.086 | 0.085 | 0.087 |
| | Mean | 0.073 | 0.110 | 0.054 | **0.0418** |

**Figure 1. An Empirical Example**

**Actual paths for 100 players** **Average simulated paths**






**bserved results (left) and the T-period ahead predictions of the three models. The results are summarized by the proportion of the choices of strategy that yields the payoff "10" or "-10" in 5 blocks of 40 trials.**

**Actual** **EWA** **LBRIAE** **REL**