Evaluating the reliance on past choices in adaptive learning models

Eldad Yechiam and Eyal Ert

Technion – Israel Institute of Technology

Eldad Yechiam, Ph.D

Email: yeldad@tx.technion.ac.il

Behavioral Science Area, Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Haifa 32000, Israel. Phone: (972) 4-8294420, Fax: (972) 4-8295688.

Forthcoming in Journal of Mathematical Psychology:

Journal homepage:

http://www.elsevier.com/wps/find/journaldescription.cws_home/622887/description?navopenmenu=-2

This research was supported in part by the Israel Science Foundation (Grant No. 244/06) and by the Max Wertheimer Minerva Center for Cognitive Studies.

Abstract:

Adaptive learning models are used to predict behavior in repeated choice tasks. Predictions can be based on previous payoffs or previous choices of the player. The current paper proposes a new method for evaluating the degree of reliance on past choices, called Equal Payoff Series Extraction (EPSE). Under this method a simulated player has the same exact choices as the player but receives equal constant payoffs from all of the alternatives. Success in predicting the next choice ahead for this simulated player therefore relies strictly on mimicry of previous choices of the actual player. This allows determining the marginal fit of predictions that are not based on the actual task payoffs. To evaluate the reliance on past choices under different models, an experiment was conducted in which 48 participants completed a three-alternative choice task in four task conditions. Two different learning rules were evaluated: An interference rule, and a decay rule. The results showed that while the predictions of the decay rule relied more on past *choices*, only the reliance on past *payoffs* was associated with improved parameter generality. Moreover, we show that the Equal Payoff Series can be used as a criterion for optimizing parameters resulting in better parameter generalizability.

Key words: Reinforcement learning; cognitive models; model selection; validity

The goal of the current paper is to present a method for evaluating adaptive learning models and for optimizing model parameters. Adaptive learning models are used to predict behavior in repeated individual and multi-player games. In these tasks, the player chooses repeatedly from multiple alternatives and receives immediate feedback after each choice without prior information concerning the alternatives' payoff distribution. Recently, there has been a movement towards modeling adaptive learning behavior at the level of the individual decision maker (Busemeyer & Stout, 2002; Busemeyer & Wang, 2000; Ho, Wang & Camerer, 2006; Erev & Barron, 2005; Haruvy & Erev, 2002; Stahl, 1996; Wallsten, Pleskac & Lezuez, 2005; Wilcox, in press; Yechiam & Busemeyer, 2005). This approach grew out of the realization that individuals are sufficiently different that pooling them together implies a grave misspecification (Estes, 1956; Haruvy & Erev, 2002; Siegler, 1987). Evaluation is usually based on the accuracy of 'next choice ahead' predictions given the previous outcomes of the player. Yet these predictions can be based on two independent factors: responses based on the previous payoffs and responses that are independent of previous payoffs and rely only on the choice history of the player. The current method evaluates the impact of the relative influence of these two factors.

Some degree of reliance on previous choices (made by the player) appears in most adaptive learning models (see Erev & Haruvy, 2005), and is due to the fact that in most models (a) the (modeled) attractiveness of an alternative can be improved by the selection of the alternative, and (b) the past selection of an alternative is associated with the past attractiveness of the alternative. In this way, the reliance on previous choices adds

additional strength to the model because past choices act as “crutches” that guide the model towards correct future choices.

Assessing the degree of reliance on previous choices is particularly important for applications of learning models to the study of cognitive processes (see e.g., Busemeyer & Stout, 2002; Cohen & Ranganath, 2005; Wallsten, Pleskac & Lezuez, 2005; Yechiam, Busemeyer, Stout & Bechara, 2005). In these investigations many data points are collected so that the estimated parameter values for each performer are hypothesized to be the same as the “population parameters” – in this case, consistent latent constructs within the individual. Moreover, in all of the above models, some latent constructs are assumed to represent the internal response style to *previous payoffs*. For example, in the Expectancy-Valence model (Busemeyer & Stout, 2002), used to model behavior in the Iowa Gambling Task (a popular task employed in clinical and neurological assessment; Bechara, Damasio, Damasio & Anderson, 1994), there are three parameters: One denoting the weighting of gains compared to losses, another denoting the weighting of recent as compared to past payoffs, and a final parameter denoting choice consistency (the consistency between beliefs based on payoffs and actual choices). All of these parameters are argued to measure consistent traits in the individual’s response to payoffs. It is therefore important to assess how much accuracy is achieved when the model actually responds to payoffs in the task; and how much is achieved strictly due to the reliance on past choices. If the degree of model accuracy is not improved by the response to previous payoffs, the estimated parameters may be meaningless in terms of the individual’s response style, and may reflect only mimicry of past choices.

In addition, the Erev-Haruvy critique (Erev & Haruvy, 2005) indicates that while the reliance on past choice can improve the accuracy of the model for next-step-ahead predictions, it also diminishes the ability to use the estimated model parameters for predicting behavior in different task conditions. The reason is that higher degree of reliance on past choices reduces the relative weight on payoff-related variables which appear to be much more useful for predicting behavior in new tasks. Evaluating the reliance on previous choices may therefore improve the ability to reliably employ the parameters for predicting behavior in different tasks.

The present paper proposes an evaluation method that distills the overall model accuracy to choice-based and payoff-based components. Secondly, we empirically assess the degree of reliance on previous choices under different learning rules and task conditions. Thirdly, we examine the effect of such reliance on the generalizability of the model's predictions to different conditions (the Erev-Haruvy critique implies a negative effect). Finally, we examine if the current evaluation method could be used for optimizing model parameters.

Theoretically, the predictions of a model are considered to be based partly on previous payoffs, partly on mimicry of previous choices, and partially on an interaction between previous choices and previous payoffs. The goal of the proposed method is to assess the part that is based strictly on previous choices without the influence of past payoffs. This part represents mere mimicry of past choices. The current method does not attempt, however, to disentangle potential interactions between past choices and past payoffs. The reason is that the payoff element in such interactions is presumably important for any potential model that aims to capture internal responses to payoffs, but

isolating this element from the choice element in the interaction is not always possible, because it is not always clear how in fact these two elements interact.

The proposed method is called Equal Payoff Series Extraction (EPSE). It uses a simulated player to assess the degree of reliance on past choices. For this simulated player the payoff series for the different alternatives are made to be equal (i.e., all of the alternatives consistently produce the exact same payoff), so that assuming no reliance on past choices the model should not be able to correctly predict the individual's future choices, compared to a random prediction. The fit of the model to the simulated individual therefore represents the accuracy of the model for predicting future choices based strictly on past choices. This component of the accuracy of the learning model (produced by the simulated player) can be deducted from the overall accuracy (produced by the actual individual player), to produce the improvement in accuracy based on past payoffs. If there is no improvement at all based on past payoffs, this implies that the model bases its prediction and parameter estimation on past choices rather than on previous payoffs.

The present investigation uses the proposed method to compare and evaluate the mimicry component of two learning rules: Delta based learning (e.g., Busemeyer & Myung, 1992; Gluck & Bower, 1988; Rumelhart & McClelland, 1986; Sarin & Vahid, 1999; Sutton & Barto, 1998) and Decay- Reinforcement learning (e.g., Erev & Roth, 1998; Yechiam & Busmeyer, 2005). The current experimental evaluation of the two learning rules employs a task in which the decision maker chooses repeatedly between a sure payoff and two riskier prospects. This general task is evaluated in eight conditions that differ from each other by: (a) the expected value of each alternative, (b) the

possibility of losses associated with the riskier alternatives, and (c) the degree of noise within each alternative. Note that the task, although specific, has the properties needed for the examination of the suggested hypotheses, as it enables the examination of both model fit and generality. Thus, it can be a starting point for appreciating whether model evaluation with the EPSE method might reveal new and important characteristics of these models and lead to a better understanding of their implications.

The remainder of the paper is organized as follows. Section 1 formally presents the current extraction method. Section 2 presents the learning rules compared here, previous findings, and the relevance of the proposed method to the evaluation. Section 3 presents a new experimental evaluation of the different learning rules. Section 4 presents the possibility of using the present method to optimize parameters for better consistency and generalizability. The discussion section summarizes the value and limitations of the EPSE method, and the implications of the results.

1. Equal Payoff Series Extraction (EPSE)

Methods used for model evaluation at the individual level often rely on optimizing the accuracy of 'one step ahead' predictions generated by each model for each individual (for an alternative approach, see Wagenmakers, Grünwald, & Steyvers, 2006). Specifically, define $Y(t)$ as a $T \times 1$ vector, representing the sequence of choices made by an individual up to and including trial T; define $X(t)$ as the corresponding sequence of payoffs produced by these choices; and define $\Pr[G_j(t+1) \mid X(t)]$ as the (predicted) probability that alternative j will be selected on trial $t+1$ by a model with a certain

parameter vector given the previous outcomes. The accuracy of this prediction for each choice trial is measured using the log likelihood criterion:

$$LL_{model} = \text{lnL (model} \mid X(t)\text{)} = \sum_t \sum_j ln(\Pr[G_j(t+1) \mid X(t)]) \cdot \delta_j(t+1) \quad (1)$$

Where the term $\delta_j(t+1)$ denotes the alternative chosen on trial $t+1$. To optimize the log likelihood for each participant and model, a parameter search is conducted (there are different methods; we use the robust method proposed by Nelder & Mead, 1965). This generates a set of solutions. The best solution is the one that maximizes the log-likelihood criterion.

The accuracy of the learning model is usually compared to a baseline model that assumes no learning. One model that can be used is a random model. Under the random model the probability of choosing alternative *j* from *k* alternatives in the next step ahead is simply 1/*k*. An alternative baseline model treats the rates as free parameters to be optimized (this so called Bernoulli model is detailed below). The final fit index is therefore a difference score obtained by comparing the log likelihood score of the learning model and the baseline model used (see Busemeyer & Wang, 2000):

$$G^2 = 2 \cdot [\, LL_{model} - LL_{baseline}] \quad (2)$$

Under the EPSE method for each individual there is a simulated player, which denotes an individual that makes the exact same set of choices for alternatives producing constant payoff series with the same magnitude. The alternatives' constant equal payoff is

calculated as the average gains and losses experienced by the actual player. This payoff magnitude is assumed to be similar enough to the actual payoffs of the individual[1].

Each model's parameters are estimated for the simulated player as well using the same comparison with the baseline model. This produces a fit score, called G'^2 for the simulated player. Finally, the fit from the actual individual is compared to the model fit for the simulated player, as follows:

$$\Gamma^2 = G^2 - G'^2 \qquad (3)$$

where Γ^2 is the corrected G^2 score without the component of the fit based merely on mimicry of previous choices G'^2; Γ^2 denotes the marginal increase in fit when the predictions are not based merely on previous choices (as in the simulated player) but also on previous payoffs (as in the actual individual). Note that Γ^2 is independent of the exact baseline model used.

2. A comparison of learning models

An examination of the learning models used in previous studies reveals that most models employ three groups of assumptions: first, a utility function is used to represent the evaluation of the payoff experienced immediately after each choice; second, a learning rule is used to form an expectancy (or propensity) for each alternative, which summarizes the experience of all past utilities produced by each alternative; third, a choice rule selects the alternative based on the comparison of the expectancies (see

[1] The robustness of the EPSE method for different payoff sizes is examined in the study below.

Yechiam & Busemeyer, 2005). Different learning rules have varying degrees of dependency on past choices in making future predictions. In the present study two learning rules that posit different assumptions about the process of expectancy updating are compared.

2.1. Utility.

The evaluation of gains and losses experienced after making a choice is represented by a prospect theory type of utility function (Kahneman & Tversky, 1979). The utility is denoted $u(t)$, and is calculated as a weighted average of gains and losses produced by the chosen alternative in trial t.

$$u(t) = W \cdot win(t)^{\gamma} - L \cdot loss(t)^{\gamma} \quad (4)$$

The term $win(t)$ is the amount of money won on trial t; the term $loss(t)$ is the amount of money lost on trial t; W and L are parameters that indicate the weights to gains and losses, respectively. For parsimony, it is assumed that $L = 1- W$ (see Yechiam et al., in press). Accordingly, a single parameter W denotes the relative weight given to gains over losses. The W parameter is constrained between 0 and 1, representing exclusive weighting to losses or gains, respectively. The parameter γ determines the curvature of the utility function. In the current study, given the small amounts of money (less than $1) earned on each trial, the value of γ was set to 1 (see also Yechiam & Busemeyer, 2005, 2006).

2.2. Updating of expectancies

Two classes of models have been proposed to account for the way new information is accumulated after making a choice (Yechiam & Busemeyer, 2005). Under one class, the decision-maker's representation of choice alternatives changes only if an alternative is selected. This class of models is labeled "interference" models, because the representation is changed by relevant events and not simply as a function of time. In a second class of models, the representation can change even if no new information about a particular alternative is presented (e.g., as a function of time). This second class of models is labeled "decay" models. The current study contrasted two models from each class that were found to have the most accurate predictions in a previous study (Yechiam & Busemeyer, 2005). A Delta learning rule was used to as an example of an interference class model, and a Decay-reinforcement model (Erev & Roth, 1998) was studied as an example of the decay class.

Delta model. Connectionist theories of learning usually employ a learning rule called the Delta learning rule (see Gluck & Bower, 1988; Rumelhart & McClelland, 1986; Sutton & Barto, 1998). It has been applied to learning in decision tasks by Busemeyer and Myung (1992) and by Sarin and Vahid (1999). The expectancy E_j for alternative j is updated as a function of its value in the previous trial (which reflects the past experience), as well as on the basis of new payoffs, as follows:

$$E_j(t) = E_j(t\text{-}1) + \phi \cdot [u(t) - E_j(t\text{-}1)] \cdot \delta_j(t) \qquad (5)$$

On each trial t the expectancy $E_j(t)$ is equal to that of the previous trials $E_j(t\text{-}1)$. In addition, if alternative j is selected in trial t, then its expectancy changes. The formula

includes a dummy variable $\delta_j(t)$ which is a weight associated with the chosen alternative. $\delta_j(t)$ equals 1 if alternative *j* is chosen on trial *t*, and 0 otherwise. This means that for all the alternatives that are not chosen, the expectancy does not get updated. When an alternative gets chosen ($\delta_j(t)$ =1), the expectancy is updated. In this case, a change occurs in the direction of the prediction error given by $[u(t) - E_j(t)]$.

The parameter ϕ is the learning rate parameter. It dictates how much of the expectancy is changed by the prediction error. The parameter is bounded between 0 and 1. In this range, the effect of a payoff on the expectancy for an alternative decreases exponentially as a function of the number of times a particular alternative was chosen. Accordingly, recently experienced payoffs have larger effects on the current expectancy as compared to payoffs that were experienced in the more distant past.

Decay-Reinforcement Rule. More recently, Erev and Roth (1998) added a decay or discount parameter to the reinforcement-learning model, which can be represented by the following equation:

$$E_j(t) = \phi \cdot E_j(t\text{-}1) + \delta_j(t) \cdot u(t) \qquad (6)$$

In this learning rule, the past expectancy is always discounted, regardless of whether an alternative is chosen and new payoff information is experienced. This is implemented by the fact that the past expectancy of all alternatives $E_j(t\text{-}1)$ is multiplied in each trial by the recency parameter ϕ (whose value is constrained to be smaller than or equal to 1). The decay formula enhances the model flexibility in mimicking past choices because it simultaneously "pushes" the previously chosen alternative (if its payoffs are positive) and

"punishes" the unchosen ones. Consequently, in the study below we used a three-alternative task that increases this difference between models compared to a binary task.

Under both models it was assumed that the initial expectancy $E_j(1)$ is equal to zero. In addition, it was assumed that the unchosen alternatives gain the average expectancy of the chosen alternative until they are chosen for the first time (for similar assumptions in games, see Erev & Roth, 1998; Harsanyi & Selten, 1998; Stahl, 1999).

2.3. Choice rule

In adaptive learning models the choice on each trial is determined by the expectancies for each alternative. We used a ratio-of-strength choice rule, which assumes that the choice made on each trial is a probabilistic function of the relative expectancies of the alternatives (Luce, 1959), as follows:

$$\Pr[G_j(t+1)] = \frac{e^{\theta \cdot E_j(t)}}{\sum_k e^{\theta \cdot E_k(t)}} \qquad (7)$$

where θ controls the sensitivity of the choice probabilities to the expectancies. Setting $\theta(t) = 0$ produces random guessing; on the other hand, as $\theta \rightarrow \infty$ we recover a strict maximizing rule. The probability of choosing the alternative producing the largest expectancy increases according to an S shaped logistic function with a slope (near zero) that increases with θ. Following Yechiam (2006), a constant choice sensitivity *c* was assumed, where $\theta = 3^{10 \cdot c} - 1$. The parameter c was limited between 0 and 1, permitting the full range between a random ($\theta \approx 0$) and highly deterministic ($\theta > 700$) choices. Increasing the bounds beyond these values does not change the results reported below.

2.4 Model evaluation

The different models were evaluated using three methods: The conventional fit index method, the EPSE method (detailed above), and an examination of parameter generalizability (Yechiam & Busemeyer, 2006).

Model fit was compared to the Bernoulli baseline model. Under the latter model the choice probabilities for each choice option are assumed to be constant and statistically independent across trials:

$$\Pr[\, G_j(t+1)] = p_j \qquad (8)$$

The parameters in this baseline model correspond to the proportions of choices pooled across all choice trials. For example, in the current three alternative tasks the estimated choice probabilities are p_1, p_2, $p_3 = 1 - p_1 - p_2$; and p_1 and p_2 are the free parameters. Therefore, a learning model can do better than the baseline model only if it explains learning effects or other trial-to-trial dependencies. The EPSE is robust to the exact baseline used as long as the predictions of the baseline model do not depend on the payoff. Still, we considered it important to determine whether in practice there is a point in using a learning model in the first place over a model that assumes no learning.

In addition to examining model fit using the traditional method and the EPSE method, we examined the generality of the different models. Yechiam and Busemeyer (2006) suggested a Generalizability test at the Individual Level (GIL). In this method the parameters estimated in one task are used to form predictions for the choices made by the same individual in another task. High GIL implies that the parameters estimated in a

specific task describe the behavior of the individual in substantially different task contexts. Low GIL implies that the parameters are highly task specific (or in other words are not useful to describe the individual's behavior in robust settings).

3. Study: Model comparison under different task conditions

A controlled experiment studied the degree of reliance on past choices under two learning rules (Delta and Decay-Reinforcement) in four variants of a multiple-choice task, described in Table 1. The task includes three alternatives, one producing Safe (constant) payoffs (S), another producing Medium risk (low variance) payoffs (M), and a third producing Risky (high variance) payoffs (R). Under one within subject condition the expected value was equal for all alternatives (S=M=R). Under another condition the expected value was higher for the riskier alternatives (S<M<R). It was expected that the move to the latter condition (S<M<R) would lead people to take more risk. However, following Yechiam and Busemeyer (2006) it was expected that despite the predicted change in risk taking, the parameters of the models would still be consistent across different individuals; and would enable to make predictions from each condition to the other condition.

It was further predicted that successful mimicry of past choices would be associated with high fit in one step ahead predictions due to the association between past and future preferences (see Haruvy & Erev, 2002), but with low generalizability at the individual level due to the smaller effect of task payoffs on the model predictions. The logic that underlies this assumption is that while reliance on prior choices is one way in which a model can improve its prediction, generality beyond a certain task emerges due

to consistency in the style of responding to outcomes rather than to the style of responding per se (see e.g., Busemeyer & Stout, 2002; Wallsten et al., 2005).

To examine the robustness of the predicted results, the task was replicated in different forms (Following Katz, 1964). Under one within-subject condition the risky alternative produced losses (LOSS condition), and under another condition a constant (of 2 points) was added to all alternatives so that the risky alternative did not produce losses (GAIN condition). Evaluating behavior in both situations is potentially interesting because individuals might apply different cognitive strategies in situations where losses are possible (Erev & Barron, 2005). Note that in the GAIN condition all of the models are reduced to two-parameter models because there is no parameter indicating the weighting of gains compared to losses. Consequently, because of the large differences between models in the LOSS and GAIN condition, we only compared the generalization between the S=M=R condition and the S<M<R condition.

Finally, as a secondary manipulation, we studied the (between-subject) effect of adding a noise factor (uniformly distributed between 0 and 1 and rounded to the closest hundredth) to the payoffs indicated in Table 1. Following Busemeyer and Townsend (1993) and Erev and Barron (2005) it was predicted that a noise factor would decrease payoff sensitivity, resulting in greater reliance on past choices than on past payoffs on each trial. However, an alternative assumption is that a relatively small noise factor might make the payoff on each trial more salient and distinct, thereby increasing the reliance on past payoffs.

3.1. Participants

Forty-eight undergraduate students from the Israel Institute of Technology (24 males and 24 females) participated in the experiment. All of the students were from the Faculty of Industrial Engineering and Management. All participants were paid in cash whatever monetary bonuses they had earned in association with their performance. Payoffs ranged from 15 NIS to 35 NIS (1 NIS = $ 4.5). Participants were randomly allocated to the two experimental (Noise) conditions, with an equal proportion of males and females in each condition.

3.2. Procedure and apparatus.

Participants were informed that they would be playing different "computerized money machines" (see a translation of the instructions in Appendix A) but received no prior information as to the game's payoff structure. Their task was to select one of the machine's three unmarked buttons in each of 100 trials. The location of three alternatives was randomized across different participants. The number of trials was unknown to the players. Payoffs were contingent upon the button chosen and were drawn from the three distributions described above. Two types of feedback immediately followed each choice: (1) The basic payoff for the choice, which appeared on the selected button for two seconds, and (2) an accumulating basic payoff counter, which was displayed constantly. At the end of each task participants were briefed as to their total accumulated bonus.

The order of the task was partially controlled and partially randomized. Half of the participants were presented with the GAIN condition before the LOSS condition and the other half were presented with the reverse order. The two expected value conditions were performed consecutively within the GAIN and LOSS conditions (e.g., GAIN-

S=M=R, GAIN-S<M<R, LOSS-S=M=R, LOSS-S<M<R), and their order was randomized.

3.3. Results

Behavioral patterns. The choice proportions under the different conditions are summarized in Figure 1. The results show that participants tended to take more risk (pick S less) in the S<M<R condition ($F(1,46) = 12.84$, $p < .01$, $MSE = .04$), in the LOSS condition ($F(1,46) = 5.90$, $p < .05$, $MSE = .08$), and in the No-noise condition ($F(1,38) = 10.22$, $p < .05$, $MSE = .10$). Moreover, a significant interaction was found between the gain/loss domain and noise ($F(1,38) = 6.89$, $p < .05$, $MSE = .08$): the tendency to take more risk in the loss domain appeared mostly in the No-noise condition. This finding appears to be consistent with other studies that suggest a tendency of decision makers to prefer alternatives that produce some degree of variance (see Sonsino, Erev & Gilat, 2006). For conciseness, post hoc analyses are not detailed here (for a replication, see Erev, Ert & Yechiam, 2006).

Robustness of the EPSE method. The fit indices for the competing models appear in Table 2. The BIC correction (Schwartz, 1978) was applied to the G^2 and G'^2 scores[2]. To examine the robustness of the current reliance on equal payoff series, different payoff magnitudes used for simulating data were compared. Recall that the original payoff magnitude was the average of the gains and losses P experienced by the player in each

[2] Specifically, in the LOSS condition the learning model has one more parameter than the Bernoulli baseline model (three compared to two). Consequently, the G^2 and G'^2 scores were penalized by $\ln(N) = \ln(100) = 4.6$, where N is the number of trials.

condition (were P is a vector including a gain component and a loss component). This was contrasted with payoff magnitudes three times higher ($3 \cdot P$) or lower ($1/3 \cdot P$) than the actually experienced gains and losses. The G'^2 scores obtained using different fixed payoff magnitudes, presented on two right most columns in Table 2, were almost identical. This indicates that the measure is stable and robust to payoff magnitude in the current task conditions.

Model comparisons. Our first analysis compared the fit indices for the two learning rules across all of the eight experimental conditions. The results show that whereas the fits of both learning models were superior to the fit of the Bernoulli baseline model (i.e., $G^2 > 0$ across all conditions), the fit of the Decay-Reinforcement model was better than the corresponding fit of the Delta model (24.7 compared to 13.4; $t(191) = 3.91$, $p < .01$). However, a larger component G'^2 from the fit of the Decay-Reinforcement model was achieved based strictly on mimicry of past choices ($t(191) = 9.23$, $p < .01$). Accordingly, the marginal increase in fit Γ^2 based on responses to payoffs was significantly better for the Delta model ($t(191) = 2.91$, $p < .01$).

Another way to represent the results is by the proportion of individuals for which $\Gamma^2 > 0$ (or $G^2 - G'^2 > 0$). In this way, under the Delta model, the marginal increase in fit ($\Gamma^2 > 0$) was larger than zero for 78.1% of the participants, compared to only 59.4% in the Decay-reinforcement model ($Z = 3.96$, $p < .01$). Namely, under the Decay-reinforcement model for a larger proportion of the participants (41%, about 100% more than in the Delta model) the predictions relied strictly on previous choices and were not improved by the addition of the actual task payoffs. These findings were replicated across

all eight conditions. As there were no differences between noise conditions, all subsequent analyses were conducted across the two noise conditions.

The second model comparison analysis examined the Generalizability at the Individual Level (GIL) of each model. In this method the parameters of the model, estimated in each expected value condition (S=M=R, S<M<R) for one step ahead predictions, were used to generate the full simulation path in the same or in the other expected value condition. In other words, this method creates multiple-step-ahead predictions of each model for each condition. One thousand simulations were generated to produce a distribution of choice sequences from a given model in the high payoff condition, and these results were averaged to produce the probability of choosing each choice option on each trial[3]. We then examined the mean square deviation of the model's predicted probability as compared to the observed proportion of choices on each trial, averaged across noise conditions and expected value conditions. We calculated the GIL as the percent of predictions better than a random prediction, using Mean Square Deviation (MSD) as a distance measure.

The results are described in Table 3 (the EPS optimization will be discussed later). First, both models produced better predictions than a random model in all conditions for the majority of the participants. Secondly, the generalization of the Delta model (-MSD in the simulation in a different condition) was significantly better than for the Decay-Reinforcement model in both the GAIN and LOSS conditions (across the two conditions, $Z = 1.80$, $p < .05$; with the prediction of the Delta model being better in 57%

[3] To the extent possible we used the exact same payoffs of the actual player in a different condition. When the payoffs experienced by the player "ran out" we used a simulation based on the payoff distributions, as described in Table 1.

of the cases). The median MSD of the Delta model was 5% lower (0.21 compared to 0.22). Therefore, although the Delta model was characterized by significantly low overall fit, it improved the generalization to different payoff conditions.

Correlates of the reliance on past choices. We also examined the contribution of the different components of the model fit, G'^2 and Γ^2, to the ability of the model to produce generalizable results. For each participant we extracted the average G'^2 and Γ^2. We then examined the Spearman correlations between the average G'^2 and Γ^2 and the fit (-MSD), across the two noise conditions (with Bonferroni correction, $\alpha = 0.05/ 4 = 0.0125$). The results appear in Table 4. The only significant results, detailed here, were in the LOSS condition. The results showed that whereas G'^2 was not associated with an improvement in fit in the generalization test (GIL), Γ^2 was associated with improved fit for both the Delta ($r = 0.27$, $p < .05$) and Decay-Reinforcement ($r = 0.25$, $p < .05$) models. This indicates that reliance on past payoffs predicted the success in the generalization tests whereas reliance on past choices did not[4].

4. Equal Payoff Series Optimization (EPSO)

The results of the current analysis suggest that the reliance on past payoffs is useful for model generalization. Especially, in both of the studied models the success in the generalization test was partially predicted by the component of fit based on responses to payoffs. A natural question, therefore, is whether the EPSE method, which was used to identify this component, would also be useful for optimizing model parameters for the same purpose. This question was examined using a prediction of one step ahead seeking

[4] In the LOSS condition decision makers might be more responsive to task payoffs and less willing to adopt a strategy based on mere choices (such as “try one then the other”, etc.).

to minimize the fit of the model (above random prediction) in the simulated equal payoff series, as follows:

$$H = \text{Max}\ \{\ [lnL\ (\text{Model}|\text{EPS}) - lnL\ (\text{Random}|\text{EPS})]\ ,\ 0\ \}\ , \qquad (9)$$

$$G^{*2} = 2\cdot[\ lnL\ (\text{Model}|\ X(t)) - H]$$

Where EPS is the Equal Payoff Series (the simulated individual), H is the advantage of a certain parameter set compared to a random model on this Equal Payoff Series, and G^{*2} is the fit of the model without the advantage H. Namely, parameters are selected based on their fit to the actual data but also based on their inability to succeed beyond a random model in predicting choices for the simulated equal payoff series. Moreover, this formula ensures that a model that is *inferior* to the random model on the simulated payoff series will not be boosted artificially.

The analysis using the G^{*2} index was conducted for the model showing more promise in terms of parameter generalizability, the Delta model. To examine the impact of this adjustment on model generalizabiliy, we used a simulation analysis as before. The results, presented in the bottom rows of table 3, showed an increase in the proportion of better than random predictions in both the GAIN and LOSS conditions. The improvement for simulating multiple trials ahead in the same payoff condition was significant ($Z = 2.79$, $p < .01$) and in the generalization to a different condition it was significant on a one sided test ($Z = 1.58$, $p < .05$). Therefore, the use of the EPS criterion for estimating

parameters improved the generalizability of the estimated Delta model parameters[5]. Surprisingly, the use of the EPS criterion also improved the ability to use the parameters for predicting multiple steps ahead in given task.

5. General discussion

The results of the study demonstrate the value of the Equal Payoff Series Extraction (EPSE) method for comparing models. The EPSE method evaluates the relative weight of past choices and outcomes in determining the predictions of an adaptive learning model. It was used to shed light on previous findings (Yechiam & Busemeyer, 2005; 2006) showing that a Decay-Reinforcement learning rule produced superior fit but poor generalizability at the individual level compared to another commonly used learning rule (Delta). The component in the fit that was based on past payoffs was significantly higher in the Delta model; and it is this specific component that was associated with the ability of the model to produce generalizable predictions.

Previous studies have been pessimistic concerning the ability to meaningfully compare learning models, mainly because of their high flexibility (see e.g., Haruvy & Erev 2002; Salmon, 2001; see also Yechiam & Busemeyer, 2005). However, extracting components based on previous choices and payoffs provides a simple way to bridge across different levels of model flexibility, by deducting the accuracy that results from success in mimicry of past choices. The EPSE method is limited however, since it overcomes only one source of model flexibility, namely, mimicry of prior choices; but it ignores other sources, particularly mimicry within the model parameters (i.e., when

[5] The advantage of the EPS optimization method in the generalizability test was replicated in a two-alternative version that includes only alternatives S and M. For conciseness, this replication is not included.

distinct parameter sets make very similar predictions on a given task). To address these diverse sources of model flexibility, Yechiam and Busemeyer (2006) suggested that the evaluation process should be based on an administration of multiple tasks to the same individual. This enables the examination of the generalizability of predictions based on model parameters across tasks. The advantage of the EPSE method is that it does not rely on the administration of multiple tasks. Moreover, it complements Yechiam and Busemeyer's (2006) method in evaluating the sources of parameter generalizbility.

A specific criticism of the use of learning model with the 'one step ahead' prediction method is the poor demonstrated ability to extract parameters using this method for simulating behavior in new tasks (Erev & Haruvy, 2005). The current analysis suggests that one possible source of this problem is the mimicry component; and accordingly, one way to overcome this problem is by the use of the Equal Payoff Series (EPS) criterion for optimizing parameters (the EPSO method). This criterion selects the appropriate parameters not only according to their success in improving the accuracy of the model in predicting the next step ahead but also for their lack of ability to fit a series of choices made by the same player in the absence of any payoff differences between alternatives. In the current study using the EPS criterion in the optimization process improved the predictive power of parameters extracted with the 'one step ahead' prediction method.

Note that for some purposes the reliance on previous choices can be beneficial. For example, as we have seen, the ability of the Decay-Reinforcement model to mimic previous choices improved its accuracy in predicting next step ahead choices. Suppose then that we have a model (Model X) that yields a really good fit but relies heavily on

past choices. Another model (model Y) yields a fit that is much worse but does not rely heavily on past choices. What model should be used? The current results suggest that the answer depends on the component of the model fit that relies on payoffs, denoted here Γ^2. A larger Γ^2 component in model Y than in model X is expected to lead to more accurate predictions in tests of generalizations to different payoff conditions in model Y. Moreover, model Y is expected to measure individual differences in parameters associated with responses to payoffs more reliabily.

The answer to the question posed above therefore depends on the researcher's goal. If the researcher's objective is to examine the model's predictions in different tasks other than the one in which the parameters were estimated, or to estimate robust parameters having to do with the response to payoffs (and this is highly important to studies of cognitive processes that seek to examine the motivational system), then extensive mimicry is not expected to be helpful. For these purposes, treating the fit of the model as a whole, without subtracting or controlling for the part associated with pure mimicry, can impair the ability to evaluate different models and to estimate model parameters.

Appendix A: The instructions for the experiment task

“Hello,

In this experiment you will play a number of different games. In each game you will operate a money machine. Each button press will lead to winning or losing a number of points (depending on the button you choose). Your goal in the experiment is to win as many points as possible. There could be differences in the number of points produced by each of the buttons. Your final bonus will be determined by the total number of points earned in the game (15 points = 1 Ag.). For your information, it is highly likely that the machine would be different for each participant. Good luck”.

References

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7-15.

Busemeyer, J. R., & Myung, I. J. (1992). An adaptive approach to human decision-making: learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General*, *121*, 177-194.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment, 14*, 253-262.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432-459.

Busemeyer, J. R., & Wang, Y. -M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44*, 171-189.

Cohen, M. X., & Ranganath, C. (2005). Behavioral and neural predictors of upcoming decisions. *Cognitive, Affective & Behavioral Neuroscience, 5*, 117-126.

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review, 112*, 912-931.

Erev, I., Ert, E., & Yechiam E. (2006). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. Mimeo.

Erev, I., & Haruvy, E. (2005). Generality, repetition, and the role of descriptive learning models. *Journal of Mathematical Psychology, 49,* 357-371.

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review, 88,* 848-881.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53,* 134-140.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 128,* 309-331.

Harsanyi, J., & Selten, R. (1988). *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press.

Haruvy, E., & Erev, I. (2002). Interpreting parameters in learning models. In R. Zwick & A. Rapoport (Eds.), *Experimental Business Research* (pp. 285-300). Kluwer Academic Publishers.

Ho, T., Wang, X., & Camerer, C. (2006). Individual differences in the EWA learning with partial payoff Information. Manuscript submitted for publication.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263-291.

Katz, L. (1964). Effects of differential monetary gain and loss on sequential two-choice behavior. *Journal of Experimental Psychology, 68,* 245-249.

Luce, R. D. (1959): *Individual choice behavior*. NY: Wiley.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7*, 308-313.

Rumelhart, D. E., McClelland, J. E., and the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Volumes 1 and 2*. Cambridge, MA: MIT Press.

Salmon, T. (2001). An evaluation of econometric models of adaptive learning. *Econometrica*, *69*, 1597-1628.

Sarin, R., & Vahid, F. (1999). Payoff assessments without probabilities: A simple dynamic model of choice. *Games and Economic Behavior, 28*, 294-309.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 5*, 461-464.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116*, 250-264.

Sonsino, D., Erev, I., & Gilat, S. (2006). On the likelihood of repeated zero-sum betting by adaptive (human) agents. Manuscript submitted for publication.

Stahl, D. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior, 16*, 303-330.

Stahl, D. O. (1999). A Horse Race Among Reinforcement Learning Models. Mimeo.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Wagenmakers, E. -J., Grünwald, P., & Steyvers, M. (2006). Accumulative prediction error and the selection of time series models. *Journal of Mathematical Psychology, 50*, 149-166.

Wallsten, T. W, Pleskac, T., & Lejuez, C. W. (2005). Modeling a sequential risk- taking task. *Psychological Review, 112,* 862-880.

Wilcox, N. T. (in press). Theories of learning in games and heterogeneity bias. *Econometrica*.

Yechiam, E. (2006). Evaluating the reliability of repeated choice tasks: A cognitive modeling approach. Manuscript submitted for publication.

Yechiam. E., & Busemeyer, J. R. (2006). Evaluating parameter consistency in learning models. Manuscript submitted for publication.

Yechiam. E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience based decision-making. *Psychonomic Bulletin and Review, 12,* 387-402.

Yechiam, E., Busemeyer, J. R., Stout, J. C., & Bechara, A. (2005). Using cognitive models to map relations between neuropsychological disorders and human decision making deficits. *Psychological Science, 16,* 973-978.

Yechiam, E., Veinott, E. S., Busemeyer, J. R., & Stout, J. C. (In press). Cognitive models for evaluating basic decision processes in clinical populations. In R. Neufeld (Ed.), *Advances in clinical cognitive science: Formal modeling and assessment of processes and symptoms*. APA Publications.

Table 1: The payoff schemes of the four experimental conditions. Each condition has three choice alternatives: S (Safe), M (Medium), and R (Risky).

Expected Value	Gain/Loss	Alternative: Payoff
Equal Expected value (S=M=R)	LOSS	S: get 0 M: 50% to win 1, 50% to lose 1 R: 50% to win 2, 50% to lose 2
Unequal Expected value (S<M<R)	LOSS	S: get 0 M: 50% to win 2, 50% to lose 1 R1: 50% to win 4, 50% to lose 2
Equal Expected value (S=M=R)	GAIN	S: win 2 M: 50% to get 1, 50% to win 3 R: 50% to get 0, 50% to win 4
Unequal Expected value (S<M<R)	GAIN	S: win 2 M: 50% to get 1, 50% to win 4 R: 50% to get 0, 50% to win 6

Table 2: Means of the fit indices for the four compared models. G^2 represents the improvement in fit for the individual (compared to the Bernoulli baseline model), G'^2 represents the improvement in fit for the simulated anti-individual, $\%\Gamma^2 > 0$ represents the proportion of individuals with $G^2 > G'^2$.

				Initial analysis			Robustness analysis	
Model	LOSS/ GAIN	Noise	EV	G^2	G'^2	$\%\Gamma^2 > 0$	G'^2 (3·P)	G'^2 (1/3·P)
Delta	LOSS	Noise	S=M=R	10.5	-1.2	83	-1.2	-1.2
			S<M<R	13.5	-2.0	75	-2.0	-2.0
		No noise	S=M=R	9.3	-6.1	83	-6.1	-6.1
			S<M<R	14.4	-5.1	83	-5.5	-5.1
	GAIN	Noise	S=M=R	15.6	3.2	83	2.7	3.2
			S<M<R	9.9	1.0	63	1.0	1.0
		No noise	S=M=R	14.7	2.1	79	2.1	2.1
			S<M<R	0.8	-9.2	75	-9.2	-9.2
Decay-Reinforcement	LOSS	Noise	S=M=R	22.3	27.4	46	27.4	27.4
			S<M<R	17.9	21.4	54	21.4	21.4
		No noise	S=M=R	9.8	14.2	58	14.2	14.2
			S<M<R	15.6	2.1	71	2.1	2.1
	GAIN	Noise	S=M=R	29.9	27.5	63	27.3	27.5
			S<M<R	23.0	21.5	50	21.5	21.5
		No noise	S=M=R	29.7	25.5	63	25.5	25.5
			S<M<R	13.7	10.5	71	10.5	10.5

Note: G^2 and G'^2 values in the LOSS condition are penalized according to the BIC criterion (by $\ln(N) = 4.6$).

Table 3: Generalizability at the individual level (GIL): Proportions of individuals with predictions (-MSD) superior to a random model on a simulation under the same or a different expected-value condition (N = 96).

Model	LOSS/ GAIN	Same condition	Different condition (GIL)
Delta	LOSS	0.59	0.73
	GAIN	0.63	0.63
Decay-Reinforcement	LOSS	0.64	0.69
	GAIN	0.54	0.58
Delta – EPS Optimization	LOSS	0.78	0.84
	GAIN	0.67	0.66

Table 4: Spearman correlations between the fit of the model in the simulation test (-MSD) and the relative accuracy based on choices G'^2 and payoffs Γ^2 from the other task (N = 96).

Model	LOSS/ GAIN	G'^2	Γ^2
Delta	LOSS	0.04	0.25*
	GAIN	0.16	0.04
Decay-Reinforcement	LOSS	-0.15	0.27*
	GAIN	0.09	0.07

* = p < .0125 (using Bonferroni adjustment)

Figure 1: Proportion of choices from the Safe (S), Medium risk (M) and Risky (R) alternatives in each of the eight experimental conditions in 100 trials.

Figure 1: Proportion of choices from the risky alternative in each of the eight experimental conditions in 100 trials.