Journal of Mathematical Psychology **42**, 266–286 (1998)
Article No. MP981214

On Learning To Become a Successful Loser: A Comparison of Alternative Abstractions of Learning Processes in the Loss Domain

Yoella Bereby-Meyer and Ido Erev

Technion – Israel Institute of Technology

One of the main difficulties in the development of descriptive models of learning in repeated choice tasks involves the abstraction of the effect of losses. The present paper explains this difficulty, summarizes its common solutions, and presents an experiment that was designed to compare the descriptive power of the specific quantifications of these solutions proposed in recent research. The experiment utilized a probability learning task. In each of the experiment's 500 trials participants were asked to predict the appearance of one of two colors. The probabilities of appearance of the colors were different but fixed during the entire experiment. The experimental manipulation involved an addition of a constant to the payoffs. The results demonstrate that learning in the loss domain can be faster than learning in the gain domain; adding a constant to the payoff matrix can affect the learning process. These results are consistent with Erev & Roth's (1996) adjustable reference point abstraction of the effect of losses, and violate all other models.

The discovery that human decision makers often violate the prescription of Savage's (1954) expected utility theory (see e.g., Kahneman & Tversky, 1979; Thaler, 1987; Camerer, 1995) has led researchers to seek descriptive alternatives for this theory and the related rationality assumption. Among the most promising alternatives is the adaptive learning approach. According to this approach people are not "natural utility maximizers," but can learn to respond to the incentive structure in an adaptive fashion in certain settings. Thus, two psychological factors have to be quantified in order to predict economic behavior: the initial decision tendencies and the adaptive learning process (Roth & Erev, 1995).[1]

This research was supported in part by grant no. 604-94-1 from the Israeli Academy of Science (to Ido Erev and Amnon Rapoport) and a grant from the USA NSF (to Al Roth and Ido Erev). The research has benefited from related research and insightful conversations with Al Roth, Bob Slonim, Joachim Meyer, Racheli Barkan, Sharon Gilat, Daniel Gopher, and Amnon Rapoport.

Please address all correspondence to Yoella Bereby-Meyer, Department of Education, Ben-Gurion University of the Negev, 84105 Beer Sheva, Israel. E-mail: yoella@bgumail.bgu.ac.il.

[1] This approach does not explain why people are not rational in one-shot experiments. Rather, it focuses on predicting their sensitivity to the economic incentive structure.

0022-2496/98 $25.00

The present research focuses on the quantification of the learning process. It addresses one of the main difficulties in the development of quantitative descriptive learning models: the abstraction of the effect of losses. Losses create a problem because experimental data support the view that choice probabilities are approximately linearly related to the ratio of accumulated reinforcements (Herrnstein, 1961; Roth & Erev, 1995). Thus, to capture this robust observation, descriptive models have to assume that choice probabilities are determined by a monotonic function of the accumulated reinforcements and this function must have strictly positive values.

Five distinct solutions to this abstraction problem have been adopted in recent attempts to develop descriptive learning models. The main goal of the current paper is to compare the descriptive power of these solutions (the specific models). We hope that a better understanding of the implications of the different solutions will facilitate a discovery of a robust approximation of learning in simple decision tasks.

The paper proceeds as follows: The next section summarizes the five available solutions to the abstraction problem. This section demonstrates that the different solutions have distinct qualitative predictions concerning the effect of the addition of constants to the payoffs in simple decision tasks. Under the assumption that changes in the payoffs do not affect the parameters, the most common solution (the assumption of an exponential response rule) typically implies no effect, while the solution proposed by Erev and Roth (that includes an adjustable reference point and the truncation of negative values) implies a nonlinear effect.

An experiment that was designed to compare the descriptive power of the distinct abstractions is then presented. It examines a binary decision under uncertainty (probability learning) task. The experiment supports the adjustable reference point and truncation solution. It shows that the qualitative predictions made by this model capture the main experimental results. In addition, this model with the original parameters (proposed by Erev and Roth, 1996) provides a good quantitative fit for the data.

The implications and some limitations of these results are discussed in the conclusions.

1. THE CHALLENGE AND ALTERNATIVE SOLUTIONS

As noted above, the difficulty in abstracting the effect of losses in descriptive models of learning is a result of the fact that in the gain domain choice probabilities appear to be almost linearly related to the ratio of accumulated payoffs. Herrnstein (1961, 1970) has discovered this phenomenon in pigeon choice data. In a certain choice task (a VI–VI schedule) Herrnstein found a matching rule: The choice probabilities matched the ratio of accumulated reinforcement. Namely, in a binary choice with two alternatives (A and B) Herrnstein observed the equality

$$P(A) = P(A)/[P(A) + P(B)] = R(A)/[R(A) + R(B)],$$

where $P(J)$ is the probability of choosing alternative J and $R(J) > 0$ is the accumulated reinforcement from these choices.

Following Herrnstein (1970) and Harley (1981), Roth and Erev (1995) used this linear relation as the basis for their quantification of the Law of Effect (Thorndike,

1898). The basic model can be summarized by the following three assumptions (Luce, 1959, shows that the distinction between the three assumptions facilitates model comparison):

A1. Initial Propensities. *At time $t=1$ (before any experience has been acquired) the DM (player)*[2] *has an initial propensity to play the kth pure strategy, given by some real number $q_k(1)$.*

A2. Updating Rule. *If the DM plays the jth pure strategy at time t and receives a non-negative payoff of x_j, then the propensity to play strategy k is updated by setting*

$$q_k(t+1)=\begin{cases} q_k(t)+x_j, & \text{if } k=j; \\ q_k(t), & \text{otherwise.} \end{cases}$$

A3. Probabilistic Response Rule. *The probability $p_k(t)$ that the kth pure strategy will be played at time t is*

$$p_k(t)=\frac{q_k(t)}{\sum q_i(t)}$$

where the sum (here and throughout the paper) is over all of the DM's pure strategies i.

So pure strategies which have been played and had success tend over time to be played with greater frequency than those which had less success. Roth and Erev (1995, and Erev & Roth, in press) observed that this linear rule provides a good description of human choice behavior in a variety of decision tasks, even without any additional parameters. It tracks behavior even when its parameters (the initial propensities) are randomly selected.

These results imply that descriptive models of learning should predict an approximately linear choice function in the gain domain. Yet, the basic linear rule cannot be used when negative payoffs are possible. Five main solutions to this difficulty, suggested in previous research, are compared in the present paper.

1.2. The Different Solutions and Quantitative Models

Each of the models presented below can be described by assumption A1 as stated above, and variants of assumptions A2 and A3. To facilitate the comparison of the different solutions we will consider the models' predictions for the decision tasks studied in the experiment described below. In each of the tasks (experimental conditions) the DM participates in 500 independent trials. In each trial the DM is asked to guess which of two mutually exclusive events L or H will occur. In all trials the probability of H is 0.7 (and the probability of L is 0.3). After each trial the DM receives an immediate feedback concerning the obtained event and his/her payoff.

The different tasks differ with respect to the obtained payoffs. In condition 4, 0 the DM earns 4 point when he or she guesses correctly and loses nothing when he

[2] Although we drop the subscript for distinct players, the models considered here apply for n-person games.

or she is wrong. The other two conditions were created by subtracting a constant from these payoffs. In condition 2, −2 the payoffs are 2 for a correct and −2 for an incorrect response, and in condition 0, −4 the DM looses 4 points when he or she is wrong and earns nothing for a correct response.

Simple decision problems of this type, referred to as probability learning tasks, were studied extensively in the 1950s and 60s. Whereas the optimal response is always to choose the most common event (guess "H"), the literature reveals that DMs are slow learners. After 100 and 200 trials they tend to "probability match;" that is, to select "H" in 70% of the trials.[3] With longer experience DMs slowly move toward the optimal choice (see Edwards, 1961). In addition, high payoffs speed the learning process (Siegel, Siegel & Andrew, 1964). The addition of constant to the payoffs was not studied by previous research.

Three specific characteristics of the current task are utilized below in the presentation and the derivation of the predictions of the distinct models. We will assume that DMs: (1) Consider only two strategies, (2) know the outcomes of both strategies after each trial, and (3) have uniform initials tendencies. In addition, we assume that the experimental manipulation cannot change the underlying learning process (the parameters of the model).

1.2.1. *A Low Reference Point Solution (the LRP Model)*

Erev and Roth (1996) proposed two variations of the basic model to address potential losses. The first variant assumes that the reinforcement is a function of the objective payoff (from choosing j) at trial t (x_j) and this function, $R(x_j)$, returns nonnegative values. Specifically, in a model referred to as the LRP model, Erev and Roth utilized the function

$$R(x_j) = x_j - X_{\min}$$

Where $X_{\min}$ is the worst possible outcome.

In addition to this modification, the LRP model includes an abstraction of two additional important characteristics of human and animal learning: Generalization (and experimentation) and Recency. These assumptions were quantified by the following generalized version of assumption A2:

$$q_k(t+1) = (1-\phi)\, q_k(t) + E_j(k, R(x_j)). \qquad (\mathrm{A2}^{\mathrm{lrp}})$$

In assumption $\mathrm{A2}^{\mathrm{lrp}}$, ϕ is a forgetting (or recency) parameter which slowly reduces the importance of past experience. $R(.)$ is the function, defined above, which translates payoffs into rewards, and E is a function which determines how the experience of playing strategy j and receiving the reward $R(x_j)$ is generalized to update each strategy k.

[3] It is important to note that the similar names do not imply that Herrnstein's "matching law" predicts "probability matching." Herrnstein's matching law, as quantified above, predicts slow learning toward the optimal choice in condition 4, 0 (and cannot be utilized to address the conditions with negative payoffs).

Erev and Roth assumed a Gaussian generalization/error function. For the binary case this function is reduced to:

$$E_j(k, R(x_j)) = \begin{cases} R(x_j)(1-\varepsilon), & \text{if } \ j=k; \\ R(x_j)\varepsilon, & \text{otherwise.} \end{cases}$$

Under the assumption of uniform initials, the LRP model has three parameters: ε and ϕ as defined above, and an initial strength parameter $S(1) = \sum q_i/$(average reinforcement from random decisions).[4]

The predictions of this model for the current tasks, with the parameters that best fit Erev and Roth's (in press) data ($\varepsilon = 0.2$, $\phi = 0.1$, and $S(1) = 9$), are presented in the left-hand column of the LRP panel in Fig. 1. (The right-hand column of Fig. 1 shows the model's predictions with estimated parameters and will be discussed below.)

The predictions (for this and all other models) were derived by running (300) computer simulations in which simulated individuals that behave according to the model's assumptions perform the three tasks described above. The predictions are summarized by the expected proportion of "H" choices in five blocks of 100 trials. As can be seen in this plot, the model predicts slow learning (in line with Edwards' findings) and no condition effect. Since $X_{\min}$ tracks the addition of constants to the payoffs, this manipulation does not affect the model's predictions.

1.2.2. An Adjustable Reference Point and Truncation Solution (the ARP Model)

Examination of the animal learning literature suggests that the effect of the reinforcements depends on a reference point that can be a function of the learner's experience. Premak (1965, 1971) proposed that reinforcements have a relative value that is determined according to a reference point. Outcomes above the reference point are perceived as reinforcements and outcomes below are perceived as punishments. The position of the reference point usually was set at zero, and outcomes greater than zero were perceived as gains and below zero as losses. However the position of the reference point could be a function of expectations, goals, and experience. A classic experiment by Tinklepaugh (1928) demonstrates the effect of experience on the reference point. Tinklepaugh taught monkeys a simple discrimination task. He reinforced one group of monkeys with bananas and the other with lettuce leaves. As long as one monkey was reinforced with bananas and the other with lettuce the task was learned quickly. However when the monkey that normally was rewarded by bananas received lettuce instead the accuracy decreased rapidly. For this monkey the lettuce leaves were perceived as punishment. The experience with the bananas as a reward set the reference point for this monkey higher than the reference point of the monkey that received lettuce from the beginning. Similar results were obtain by Tolman (1932) in an experiment with rats. To address these results Erev and

[4] In the current setting players are assumed to know the possible (average and minimal) payoffs. The model can approximate behavior even when these values are not known under the assumption that they can be learnt in the first few trials.

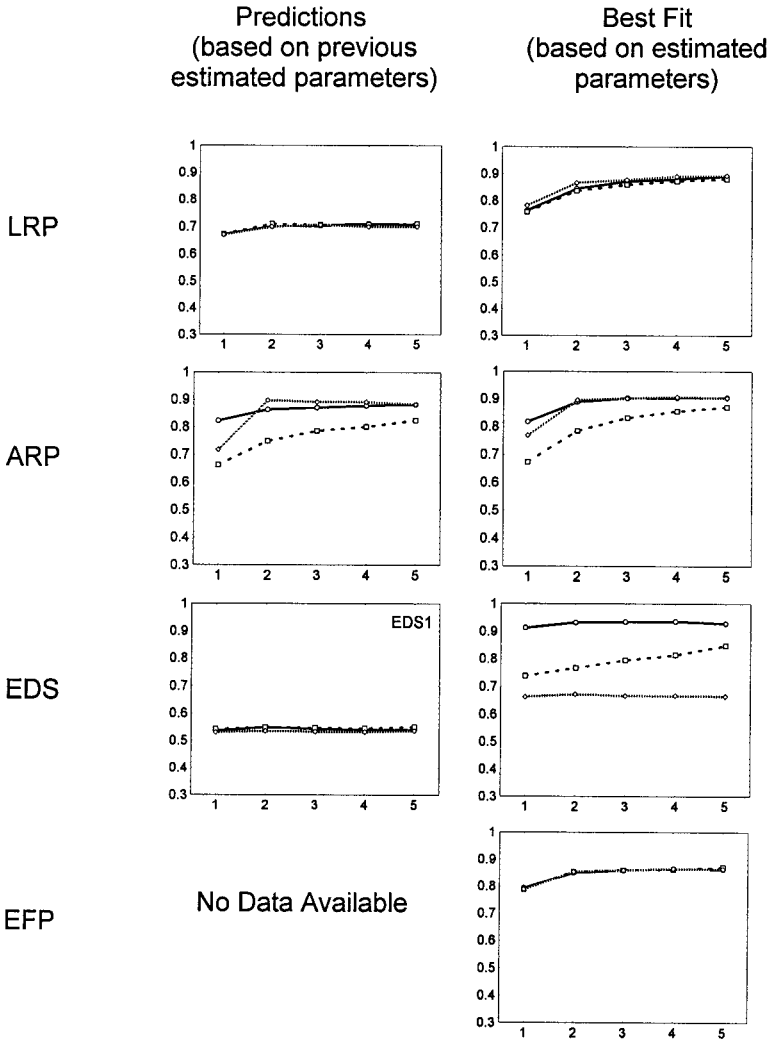

FIG. 1. The models' predictions (with parameters that were estimated in previous studies) and best fit (with estimated parameters that minimize the MSD scores). Each curve shows the proportion of "H" choices in 5 blocks of 100 trials in one of the conditions.

Roth explored a family of models with adjustable reference points (the ARP models). These models that generalize the LRP model assume that the reinforcement function changes with time. Specifically,

$$R_t(x_j) = x_j - \rho(t)$$

where $\rho(t)$ is the reference point in trial t. The reference point at the beginning of the experiment is denoted by $\rho(1)$, and it is assumed to be updated by the following linear weighting function:

$$\rho(t+1) = \begin{cases} (1-w^+)\,\rho(t) + (w^+)\,x_j & \text{if} \quad x_j \geqslant \rho(t), \\ (1-w^-)\,\rho(t) + (w^-)\,x_j & \text{if} \quad x_j < \rho(t), \end{cases}$$

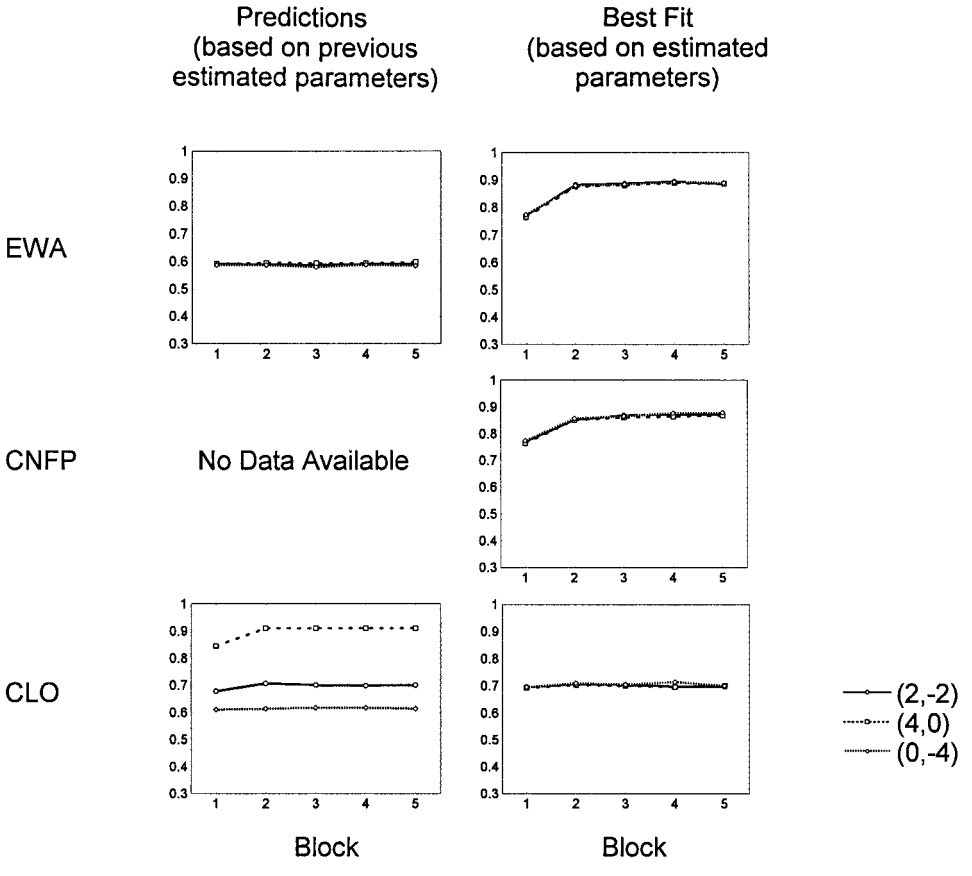

FIG. 1—*Continued.*

where w^+ and w^- are the weights assigned to positive and negative reinforcements, respectively.

Finally, to address the problem of negative reinforcements a truncation rule is added. Thus A2 is replace with:

$$q_k(t+1) = \max[\upsilon, (1-\phi)\, q_k(t) + E_j(k, R_t(x_j))]. \qquad (\mathrm{A2}^{\mathrm{arp}})$$

where $\upsilon > 0$ is a technical parameter which ensures that propensities remain positive.

After reducing the number of initial propensities parameters (by the uniform initial assumption explained above) the ARP model has seven parameters (ε, υ, ϕ, $S(1)$, $\rho(1)$, w^+, and w^-).

The predictions of the model with the parameters estimated by Erev & Roth (1996, and utilized by Rapoport *et al.*, 1997; Rapoport *et al.*, in press; Erev & Rapoport, in press) are presented in the left-hand column of Fig. 1. The parameters' values are $\varepsilon = 0.2$, $\upsilon = 0.0001$, $\phi = 0.001$, $S(1) = 3$, $\rho(1) = 0$, $w^+ = 0.01$, and $w^- = 0.02$. The model predicts a nonlinear effect of a subtraction of constants. The first subtraction (from 4, 0 to 2, -2) speeds the learning process, whereas the second (from 2, -2 to 0, -4) slows it slightly. Yet the 0, -4 condition is predicted to lead to faster

learning than the 4, 0 condition. A sensitivity analysis (that will be reported in details below) shows that these qualitative predictions are robust to changes in the model parameters.

1.2.3. *An Exponential Response Rule Solution*

Whereas the two solutions considered above are based on a transformation of the accumulated reinforcements, a more popular solution is based on a transformation of the response rule. In particular, many researchers (e.g., Busemeyer & Myung, 1992; Camerer & Ho, in press; Fudenberg & Levine, 1995; McKelvey & Palfrey, 1995; Mookherjee & Sopher, 1997) assume an exponential response rule. These models replace assumption A3 above with:

$$p_k(t) = \exp[\lambda q_k(t)] \Big/ \sum \exp[\lambda q_i(t)]. \qquad (A3^e)$$

Note that the parameter λ determines the extremeness of the predictions. With high λ the model's predictions are close to being deterministic, and with low λ they move toward uniform predictions. Exponential models are more convenient mathematically as they eliminate the need for a nonmonotonic truncation function. Three types of exponential models will be considered here.

An Exponential Discounted Sum (EDS) Model. Tang (1996, and Chen & Tang, 1996) examined a model that utilizes A1, a variant of $A2^{Irp}$ that implies a discounted sum ($\varepsilon = 0$, $R(x_j) = x_j$), and $A3^e$.

Tang studied games with only non-negative payoffs and found that although the EDS model has an additional parameter, it does not lead to significantly better predictions than the basic model (with A3 linear response rule). The predictions of the EDS model with the parameters estimated by Tang ($\lambda = 0.02$, $\phi = 0.1$)[5] are presented in Fig. 1. The plot reveals very slow learning and no condition effect. Similar predictions are also made based on the parameters estimated by Chen and Tang ($\lambda = 0.006$, $\phi = 0.2$). Yet, faster learning and a clear effect for the addition of constant is predicted by the EDS given certain parameters. In particular, simulations that were run with larger λ (e.g., 0.8, 1) and small ϕ (e.g., 0.001, 0.01) reveal that with these values the model predicts fast learning around 0 (in condition 2, -2) and particularly slow learning in the loss domain (condition 0, -4).

An Exponential Fictitious Play (EFP) Model. Fudenberg and Levine (1995) studied the convergence properties of learning rules that assume a continuous and smooth best reply response rule. As an example of a function of this type they proposed an exponential fictitious play (EFP) model. This model utilizes the $A3^e$ response rule and assumes that the propensities are the expected rewards under the assumption of a static environment. That is, the DM is assumed to be a naive statistician who tries to assess the expected reward of the different alternatives. Whereas these assessments are fictitious in games (as the other players change their acts and the environment

[5] Whereas Erev & Roth chose the parameters that "best" reproduce the complete learning curve, Tang and Chen selected the parameters that "best" predict trial $t+1$, given the model and the first t trials.

is not static) they are accurate in the present one person static task. Following Camerer and Ho (in press) this model can be summarized by the following updating rule:

$$q_k(t+1) = [(N(1)+t-1)\cdot q_k(t) + v_k]/[N(1)+t], \qquad (\mathrm{A2}^{\mathrm{efp}})$$

where v_k is the payoff that the player would receive for a choice of strategy k in trial t ($v_k = x_j$ if k was chosen),[6] and $N(1)$ is a parameter that determines the weight of the initial (uniform) tendencies.

It is easy to see that this model predicts a convergence to a fixed choice probability $P(\text{“H”}) = 1/[1+\exp(-\lambda\cdot 1.6)]$; that is, it is not affected by the addition of a constant to the payoffs. The speed of convergence is expected to be a function of initial weight parameter. As the initial weight increases, the speed of convergence decreases.

An Experience Weighted Attractions (EWA) Model. Camerer and Ho (in press, 1996) developed an experienced weighted attraction (EWA) model that generalizes reinforcement learning and best reply rules. Their model can be summarized by $\mathrm{A3}^{\mathrm{e}}$ and the following variant of A2:

$$q_k(t+1) = \{(1-\phi)\cdot N(t)\cdot q_k(t) + [\delta + (1-\delta)\cdot I(t,k)]\cdot v_k\}/[\eta\cdot N(t)+1]. \qquad (\mathrm{A2}^{\mathrm{ewa}})$$

where ϕ is a forgetting parameter, $N(t) = \eta\cdot N(t-1)+1$ (for $t>1$) is a function of the number of trials the DM experienced, δ is a parameter that determines the relative effect of reinforcements and best reply considerations, $I(t,k)$ is an index function that returns the value 1 if strategy k was selected in trial t, and 0 otherwise.

Under the assumption of uniform initial propensities the EWA model has five parameters: λ as in the other exponential models considered above, initial weight $N(1)$, a forgetting parameter ϕ, an experience depreciation parameter η, and the reinforcement/expectation (imagination) parameter δ. Note that with $N(1)=1$, $\eta=0$, $\delta=0$, it coincides with the EDS model. With $\phi=0$, $\eta=1$, $\delta=1$, it coincides with the EFP model.

The predictions of the EWA model with the parameters estimated by Camerer and Ho (1996) for a 6×6 constant sum game ($\lambda=0.27$, $N(1)=12.43$, $\phi=0$, $\eta=0.94$ and $\delta=0.79$) are presented in Fig. 1. With these parameters, the model predicts very slow learning in the current tasks and no condition effect. Since the EWA generalizes the EDS and the EFP models, its sensitivity to the addition of a constant is parameter specific. Within certain parameters, the pattern implied by the EDS model is predicted here.

1.2.4. A Cumulative Normal Response Rule Solution (and the CNFP Model)

An alternative transformation of the response rule was studied by Cheung and Friedman (1994, 1996). Like Fudenberg and Levine their model is a probabilistic best reply rule. The propensity to select a certain choice is a weighted average of

[6] For example, in condition 4, 0 $v_k = 4$ if k was the “accurate” response and 0 otherwise.

the strategy's past return. In the 1996 paper they used a one parameter updating rule that can be approximated by

$$q_k(t+1) = [(1-\phi)\cdot(N(1)+t-1)\cdot q_k(t) + v_k]/[(1-\phi)\cdot(N(1)+t-1)+1]. \qquad (A2^{cnfp})$$

To solve the problem of potential negative propensities Cheung and Friedman replaced A3 with an accumulated normal response rule. For the binary case their rule implies:

$$p_k(t) = F\{\alpha + \beta\cdot[q_k(t) - q_i(t)]\}, \qquad (A3^{cnfp})$$

where F is the standard normal commutative function, α is a parameter that reflects an a priori preference for strategy k, β is a responsiveness to learning parameter, and $j \neq k$.

For the current task in which the DM has no prior information about the strategies it is natural to assume that $\alpha = 0$. Under this assumption (and an assumption of a small ϕ value), the model predicts convergence to $P(\text{“H”}) = F\{1.6\cdot\beta\}$ in all three conditions considered here. As in the case of the EFP model the speed of convergence is expected to be a function of the initial propensities parameter. With high ϕ values the model predictions move toward uniform predictions.

1.2.5. *Relative Reinforcement Solutions and a Cardinal Linear Operator (CLO) Model (March, 1996)*

In its general form the linear operator model (Bush & Mosteller, 1955) allows for outcome specific parameters. These outcome parameters determine the direction and the magnitude of linear operations on the propensities (that are equal to the choice probabilities in this model).

Whereas this solution is convenient for modeling behavior in settings in which there is little reason to assume a specific quantitative relation among the outcomes, it appears to be too weak when the payoffs are small monetary prizes. To address human decision making in gambles, March (1996) considered the following variant of the linear operator updating rule (for the two strategy, two outcome case).[7]

$$q_k(t+1) = \begin{cases} 1-(1-\phi)^{a(x_j)}\cdot[1-q_j(t)], & \text{if } k=j \text{ and } x_j \geqslant 0; \\ (1-\phi)^{a(x_j)}\cdot q_j(t), & \text{if } k=j \text{ and } x_j < 0; \\ 1-q_j(t+1), & \text{if } k \neq j; \end{cases} \qquad (A2^{clo})$$

where $a(x_j)$ returns the absolute value of x_j.

The predictions of this model for the current tasks with the parameter chosen by March ($\phi = 0.1$) are presented in Fig. 1. The main prediction is a condition effect

[7] It should be noted that March suggested this specific rule to demonstrate a general point (that reinforcement learning can lead DM to risk aversion in the gain domain and loss aversion in the loss domain). He studied two additional rules that cannot be utilized in the current setting.

TABLE 1

A Summary of the Different Solutions and Models

Model	Solution: A2: The updating rule: $q_k(t+1) =$ (when j was chosen at t and the payoff was x)	Solution: A3: the choice rule: $p_k(t) =$	Predicted effect for adding constants
Basic	$q_k(t)+x$, if $j=k$ $q_k(t)$, otherwise	$\frac{q_k(t)}{\sum q_i(t)}$	Yes
A low reference point:			
LRP	$(1-\phi)\cdot q_k(t)+E_j(k, R(x))$.	$\frac{q_k(t)}{\sum q_i(t)}$	No
An adjustable reference point with truncation:			
ARP	$\mathrm{Max}[v, (1-\phi)\cdot q_k(t)+E_j(k, R_t(x))]$	$\frac{q_k(t)}{\sum q_i(t)}$	Yes, faster learning in 2, -2 and slowest in 4, 0
Exponential response rule:			
EDS:	$(1-\phi)\cdot q_k(t)+x$, if $j=k$ $(1-\phi)\cdot q_k(t)$, otherwise	$\frac{\exp[\lambda\cdot q_k(t)]}{\sum \exp[\lambda\cdot q_i(t)]}$	Yes, but weak given the estimated parameters
EFP	$\frac{(N(1)+t-1)\cdot q_k(t)+v_k}{N(1)+t}$	$\frac{\exp[\lambda\cdot q_k(t)]}{\sum \exp[\lambda\cdot q_i(t)]}$	No
EWA	$\frac{(1-\phi)\cdot N(t)\cdot q_k(t)+[\delta+(1-\delta)\cdot I(t,k)]\cdot v_k}{\eta N(t)+1}$	$\frac{\exp[\lambda\cdot q_k(t)]}{\sum \exp[\lambda\cdot q_i(t)]}$	Yes, but weak given the estimated parameters
Cumulative normal response rule:			
CNFP	$\frac{(1-\phi)\cdot(N(1)+t-1)\cdot q_k(t)+v_k}{(1-\phi)\cdot(N(1)+t-1)+1}$	$F\{\alpha+\beta[q_k(t)-q_i(t)]\}$	No
Relative reinforcement function:			
Clo	$1-(1-\phi)^{a(x_j)}\cdot[1-q_i(t)]$ if $k=j, x_i\geq 0$ $(1-\phi)^{a(x_j)}\cdot q_i(t)$ if $k=j, x_i<0$ $1-q_j(t+1)$, if $k\neq j$	$q_k(t)$	Yes, faster learning in 4, 0

that contradicts the prediction of the ARP model: faster learning in the 4, 0 condition than in the other two conditions.[8]

1.3. Summary

The seven models presented above are summarized in Table 1. This table shows that, whereas the models are rather similar (predict slow probabilistic adjustment to payoffs), they have distinct predictions with regard to the effect of the addition of constants to the payoffs. Three models (LRP, EFP, and CNFP) predict no effect. For the other models we obtained parametric specific predictions. With the parameters estimated in previous research, the ARP model predicts the slowest learning in the gain domain, the CLO model predicts slow learning in the loss domain, and the EDS and EWA predict small differences. The experiment presented below was designed to compare the predictions with empirical results.[9]

2. EXPERIMENT

2.1. Method

Participants. Forty-two Technion students served as paid participants in the experiment. They were randomly assigned to the experimental conditions. The exact payoffs were contingent on performance and ranged from 24–26 Shekels ($8–8.5).

Apparatus and Procedure. The experiment was programmed and run using Visual Basic 3 for Windows 3.1. This system was installed on a 486PC, with a Super VGA 14″ screen.

The experiment was run for 500 trials. In line with previous probability learning experiments the decision problem was described to the participants as a prediction task. In each trial they were asked to predict the appearance of one of two colors (Blue or Red). The participants were told that their payoff will be G for a correct prediction and B for incorrect prediction. The value of G and B defined the experimental conditions.

The display consisted of three white fields. The upper field showed the cumulative score, the middle field showed the participant's prediction and the lower field showed the actual result of the sampling process.

The two bottom fields (prediction and result) were empty at the beginning of each trial. The participant's response (⟨A⟩ for Blue and ⟨S⟩ for Red) filled the prediction field. Two seconds after the response the result field was filled with a randomly selected color, the cumulative payoff field was updated in accordance

[8] Similar predictions are also made by another variant of the Bush and Mosteller model, proposed by Borgers and Sarin (1995). Borgers and Sarin assume an adjustable reference point (similar to the ARP model), and distinct forces above and below the reference point.

[9] It should be emphasized that we study the predictions of the different models and not the predictions of the different abstractions of the effect of losses. The models' predictions are a function of the abstractions of the effect of losses and other properties of the models.

to the correctness of the choice. This screen was presented for three seconds, then the colors were erased from the fields and the next trial began.

Participants were divided randomly into the three conditions (14 participants in each condition). As noted above, the different conditions varied with respect to the obtained payoffs. In condition 4, 0 the payoff for inaccurate predictions (B) was 0, and the payoff for accurate predictions (G) was 4. In addition, the participants received 2000 initial points.

Condition 2, -2 was created by subtracting 2 points from each outcome. Thus, the payoffs were $B=-2$, $G=2$. The subtracted points ($2x500=1000$) were added to the initial endowment (which was set to 3000) to insure identical objective incentives in the different conditions.[10] In condition 0, -4 the payoffs were $B=-4$, $G=0$ and the initial endowment was 4000 points. The value of each point was 0.01 Shekels ($0.003) in all three conditions.

For each participant, one of the two colors (Red or Blue) was selected to be the "high probability" accurate response. This color was the accurate response in 70% of the 500 trials. The order of the appearance of colors was randomized independently for each participant across the 500 trials.

2.2. Results

The experimental results are summarized in Fig. 2 by the proportion of "H" choices ("optimal" choices) in five block of 100 trials in each condition.

A two-ways repeated measure ANOVA (with block as the repeated measure) on the choice of the dominant color (P("H")) was conducted. The independent variables were the block and the reward condition (0, -4; 2, -2; 4, 0). The analysis revealed a significant effect of the reward condition $F(2, 39)=7.8$, $p<0.001$, and of the block, $F(4, 156)=25.31$, $p<0.0005$. Post hoc comparisons revealed a difference between the conditions 2, -2 and 4, 0 ($p<0.0006$), and between 0, -4 and 4, 0 ($p<0.005$). There was no difference between conditions 2, -2 and -4, 0.

To evaluate the robustness of the reward condition effect, we also conducted a nonparametric ANOVA (Kruskal–Wallis test) on the average choice probability (across the 500 trials). In line with the results of the traditional test, reported above, this test reveals a significant effect $H(2, N=42)=12.084$, $p<0.002$.

These significant trends favor the ARP model over the alternative models. This model correctly predicted the qualitative effect of the addition (subtraction) of constants from the payoffs.

2.2.1. Quantitative Model Comparison

Predictive Power. The left part of Table 2 presents two quantitative measures of the accuracy of the predictions made by the models, based on parameters estimated in previous studies. As explained above, estimated parameters were available

[10] That is, each choice pattern (number of "H" choices) over the 500 trials yields the same expected total payoff in the different conditions.

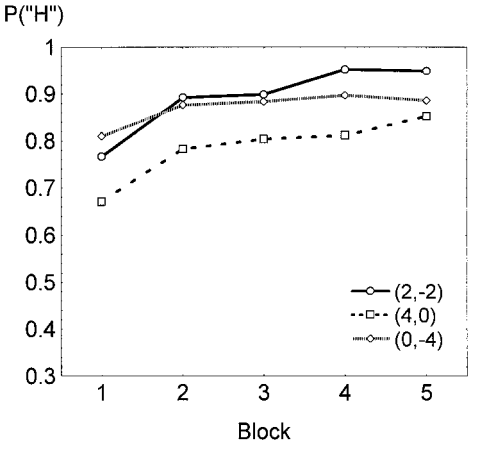

FIG. 2. Aggregated experimental results. Proportion of "H" choices in 5 blocks of 100 trials in each of the three experimental conditions.

for five of the models. The two scores are the correlation and the mean squared deviation (MSD). Each MSD score is an average of 15 (5 blocks $\times$ 3 conditions) squared deviations between the model's predictions (left-hand column in Fig. 1) and the experimental data (Fig. 2). (These values were multiplied by 100.) This panel supports the qualitative analysis: The ARP model provides the most accurate prediction of the data.

Descriptive Power. The right-hand column of Fig. 1 presents the predictions of each model with estimated parameters that minimize the MSD scores. The values of the estimated parameters and the fit of the estimated models are summarized in the right part of Table 2. Whereas a comparison of estimated models with different numbers of free parameters is not trivial, a clear picture emerges here: Even with fitted parameters, none of the alternative models outperforms the ARP model with the original parameters. Moreover, even with estimated parameters no one of the alternative models captures the observed condition effect.

2.2.2. Between-Subject Variability

The left-hand column in Fig. 3 presents the learning curve (proportion of "H" choices in 5 blocks of 100 trials) of each of the participants in the experiment (14 subjects in each condition) for each of the probability learning tasks. The middle and the right columns present samples of 14 virtual subjects from the ARP and the EWA models (that behave according to the model with the parameters that provide the best fit to the data with estimated parameters). The inspection of this figure reveals large between-subject differences in the data. The variance predicted by the ARP model is closer to the data than the EWA variance, but it is still too small. It seems that the assumption that all participants have the same parameters has to be relaxed to account for the observed variance.

TABLE 2

Mean Squared Distance (MSD, multiplied by 100) and Correlation (r) Coefficients between the Model's Predictions and the Experimental Results with Parameters That Were Estimated in Previous Studies and with Estimated Parameters That Minimize the MSD Scores

Model	Previous studies' parameters	r	MSD	Estimated parameters	r	MSD
LRP	Erev & Roth (1997) $\varepsilon=0.2,\ \phi=0.1,\ S(1)=9$	0.61	2.67	$\varepsilon=0.05,\ \phi=0.003,\ S(1)=5$	0.67	0.26
ARP	Erev & Roth (1996) $\varepsilon=0.2,\ \phi=0.001,\ S(1)=3$ $w^+=0.01,\ w^-=0.02,$ $\rho(1)=0,\ v=0.0001$	0.86	0.17	$\varepsilon=0.25,\ \phi=0.001,\ S(1)=5,$ $w^+=0.028,\ w^-=0.04,$ $\rho(1)=0,\ v=0.0001$	0.92	0.08
EDS1	Tang (1996) $\lambda=0.02,\ \phi=0.1$	−0.17	10	$\lambda=1.05,\ \phi=0.12$	0.25	1.62
EFP	No Data Available			$\lambda=1.2,\ N(1)=10$	0.7	0.31
EWA	Carmerer & Ho (1996) (The 6*6 game) $\phi=0,\ \delta=0.79,\ N(1)=12.43,$ $\eta=0.94,\ \lambda=0.27$	0.67	1.3	$\phi=0.025,\ \delta=1,$ $N(1)=12.5,\ \eta=0.86,$ $\lambda=0.24$	0.73	0.26
CNFP	No Data Available			$\phi=0.00001,\ N(1)=20,$ $\beta=0.75$	0.76	0.25
CLO	March (1996) $\phi=0.1$	−0.48	4	$\phi=0.4$	0.49	2.69

2.2.3. A Model-Based Analysis of the Robustness of the Condition Effect

Since the models considered here are probabilistic, a prediction of no condition effect on the mean choice probabilities does not imply that there will be no difference between small samples of subjects in the different conditions. Thus, in theory, the significant condition effect reported above could be observed even if one of the models that predict no effect (on the means) is correct. A two-step analysis was conducted to evaluate the likelihood of this claim. First, we simulated 20 samples of 14 simulations for each payoff condition under each model (with Table 2's estimated parameters). The 20 samples in each condition can be combined in 8000 ((20)(20)(20)) ways to create a virtual replication of the experiment. Each of the 8000 replications was then subjected to an analysis of variance of the type performed on the original experiment. The results of this analysis reveal that for 5 of the models (LRP, EFP, EWP, CNFP, and CLO) the proportion of replications with a significant condition effect is below 10%. A different pattern is observed for the ARP and the EDS models that yield a significant condition effect in all cases. The direction of the effect agrees with the experimental results in the case of the ARP model, and contradicts the results in the case of the EDS model. These results support the suggestion that the effect of the addition of a constant is not likely to be the result of random variance.

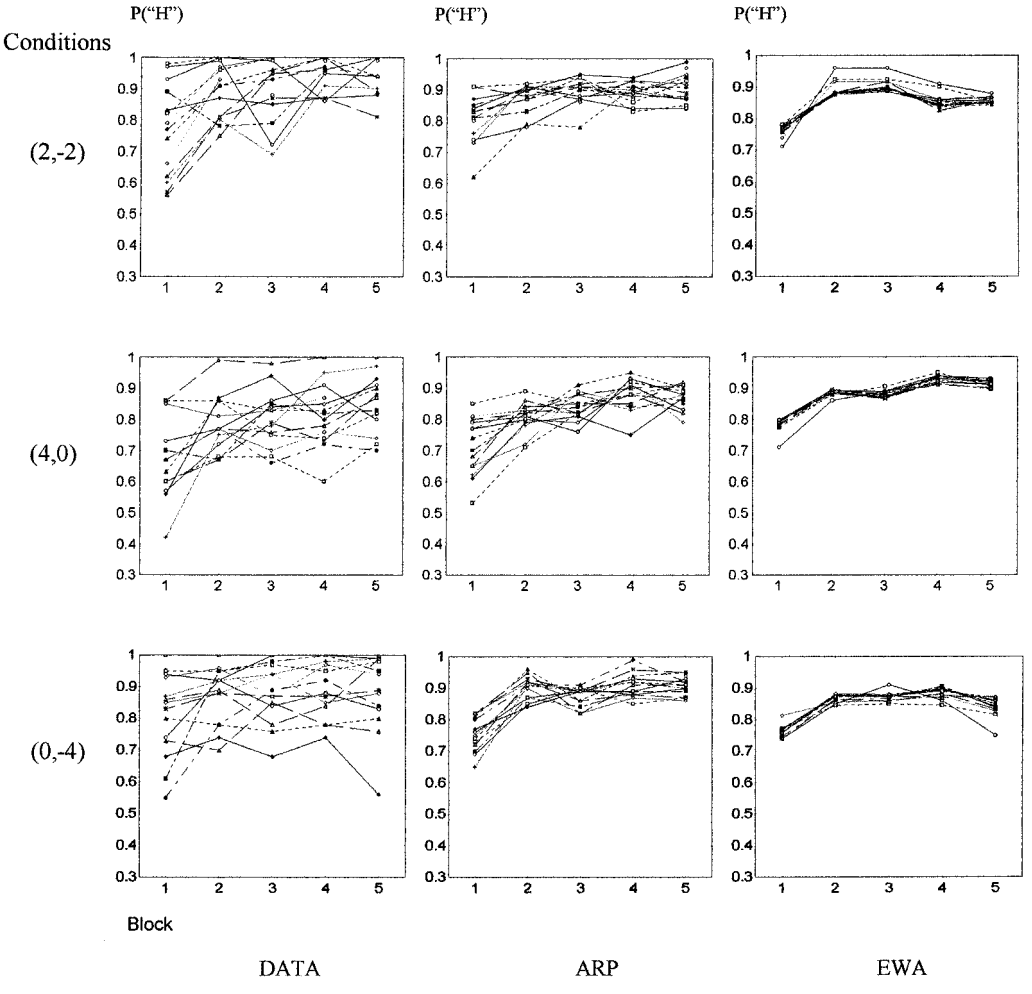

FIG. 3. Individual learning curves of the experimental participants (14 in each condition) and randomly selected virtual participants (14 in each condition) that behave according to the ARP and the EWA model with estimated parameters.

2.2.5. Sensitivity Analysis

A sensitivity analysis was conducted to evaluate the robustness of the predictions of the ARP model to the choice of parameters. Starting with Erev and Roth's (1996) parameters, the analysis examined the effect of increasing or decreasing the value of each parameter by 50%. The analysis revealed that the predicted qualitative condition effect is robust. Slower learning in condition 4, 0 is predicted with all the studied parameter values. In addition, in all cases the models' quantitative fit remains high with MSD scores below 0.3.

3. DISCUSSION

The current results demonstrate that in simple repeated decision tasks the addition of a constant to the payoffs can affect the speed of learning. In line with the results of previous probability learning studies (see review in Luce & Suppes, 1965) an extremely slow learning process was observed when the payoffs did not include negative values. The subtraction of a constant from all payoffs (that did not affect the objective incentives but implied that suboptimal choices will lead to losses) facilitated the learning speed.

Only one of the five solutions of the abstraction of losses, considered here, captures these results. According to this solution that was implemented by the ARP model, reinforcements are evaluated relative to an adjustable reference point.

The advantage of the ARP solution does not imply that the other four solutions are not valid. Models that use other solutions can account for the current results under the assumption that the models' parameters can be affected by the payoffs, or with certain additional assumptions (e.g., reinforcement functions with the characteristics of Prospect Theory's (Kahneman & Tversky, 1979) value function). The current results simply imply that, under the assumption that the parameters are not affected by the payoffs, the ARP model provides a better account for the effect of the addition of a constant than the other six models proposed in various studies and studied here.

3.1. On the Generality of the Current Results

Although the current results are sufficient to reject models that assume a learning process that is insensitive to the addition of payoffs, they do not imply that the addition of constants will always have an effect. In fact, experimental results and simple thought experiments indicate that the optimal abstraction of the effect of losses (and the addition of constants) is likely to be more complex than the abstraction implied by the ARP model. To improve our understanding of this abstraction problem it is useful to consider settings in which the predictions of the ARP model appear to hold, as well as settings in which the model is likely to fail.

3.1.1. Consistent Evidence

Previous Probability Learning Studies. Whereas previous probability learning experiments did not study the effect of the addition of a constant to the payoff directly, the effect observed here is consistent with the pattern of results observed in that research. Most importantly, Siegel and Goldstein (1959) found a strong payoff effect. They studied a probability learning task considered here (with $P(H)=0.7$) and compared a 5, 0 ($G=5$, $B=0$) condition with a 5, -5 condition ($G=5$, $B=-5$). Their results indicate significantly faster learning in the 5, -5 condition. They hypothesized that this effect is a result of the magnitude of the difference between G and B and proposed a subjective expected utility explanation. However, a subsequent study (Myers *et al.*, 1963) failed to support this explanation. Myers *et al.* found that a

multiplication of all payoffs by 10 has only a small effect on the learning speed.[11] In a recent study, Bereby-Meyer (1997) shows that both Siegel and Goldstein's and Myers' *et al.* results are predicted by the ARP model. In fact, Bereby-Meyer shows that this model (with Erev & Roth's parameters) provides a good quantitative fit (better than the fit found here) to all probability learning studies summarized in Luce and Suppes (1965).

Signal Detection Tasks. An additional demonstration of the significance and predictability of the effect of the addition of a constant to the payoffs is provided by Barkan, Zohar, & Erev (in press). This study examines binary decisions in a probabilistic signal detection task. In each trial the DM saw a numerical signal that represented a height (that was sampled from one of two distributions with similar standard deviations but different means). In line with the predictions of the ARP model,[12] Barkan *et al.* found a significant effect for the addition of constant to the payoffs.

3.1.2. Limitations

The ARP model with the parameters utilized here and in Erev & Roth (1996) implies that the initial reference point is at zero, and that the adjustment process is rather slow. Whereas these quantitative assumptions were found to provide a good fit to learning in a wide set of situations, it seems that they are not likely to hold when the DM receives clear information which implies that zero is not the appropriate reference point. Two situations of this type were considered in the literature.

Learning among Good (or Bad) Alternatives. Think about a decision setting in which all alternatives always lead to positive outcomes. For example, assume that the current task was played with $G = 104$ and $B = 100$. With the current parameters, the ARP model predicts an extremely slow learning process. Even after 500 trials the B outcome will be reinforcing and the proportion of optimal choices will be below 70%. We do not believe that this prediction is likely to hold. In fact, some results (Bereby-Meyer, 1997) show that in less extreme cases ($G = 6$ and $B = 2$; $G = -2$ and $B = -6$) the prediction of the ARP model with the current parameters is inaccurate. It seems that in these tasks, DMs quickly learn to treat G as a positive reinforcement and B as a negative reinforcement.

The Effect of Other Players' Payoff. Erev and Rapoport (in press, and see a related phenomenon in Bolton, 1991) noticed that in certain settings DMs behave as if they use other players' outcomes as their reference point. That is, a profit of 4 points might be a positive reinforcement if other players earn less, but not if other players earn more.

[11] See Mookherjee and Sopher (1997) and Slonim and Roth (1997) for a recent examination of the effect of payoff magnitude on learning.

[12] The extension of the ARP model to signal detection task is provided in Erev *et al.* (1995) and is evaluated in Erev (in press).

3.2. Conclusions

The current results clearly show that human learning can be affected by the addition of constants to the payoffs. It seems that the distinction between gains and losses affects learning. Thus, modeling this distinction can improve the potential descriptive power of adaptive learning models.

Of the five abstractions of the effect of losses considered here, the assumption of an adjustable reference point faired best. Yet it seems that a more careful quantification of the reference point is needed. Whereas the quantification proposed by Erev and Roth provides a good fit for the current data and to previous probability learning and signal detection experiments, it is not likely to be generally accurate. It seems that in some settings behavior is better approximated under the assumption of a reference point that is initially (or quickly adjusted to be) larger than the worst possible outcome. Future research is needed to provide a better understanding of the optimal quantification.

REFERENCES

Barkan, R., Zohar, D., & Erev, I. (in press). Accidents and decision making under uncertainty: A comparison of four models. *Organizational Behavior and Human Decision Processes*.

Bereby-Meyer, Y. (1997). *On the relative value of reinforcements*: *The effect of pay-off framing on learning in binary choice tasks*. Ph. D. dissertation, Technion.

Bolton, G. (1991). A comparative model of bargaining: Theory and evidence. *American Economic Review*, **81**, 1096–1136.

Borgers, T., & Sarin, R. (1995). *Naive reinforcement learning with endogenous aspirations*. Mimeo, University College, London.

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

Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning*. New York: Wiley.

Camerer, C. (1995). Choice under uncertainty. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of Experimental Economics*. Princeton, NJ: Princeton Univ. Press.

Camerer, C. F., & Ho, T. (in press). EWA learning in games: Preliminary estimates from weak-link games. In D. Budescu, I. Erev, & R. Zwick (Eds.), *Games and human behavior: Essays in honor of Amnon Rapoport*.

Camerer, C. F., & Ho, T. (1996). *Experience weighted attraction learning in games: A unified approach*. Working paper, Cal. Tech.

Chen, Y., & Tang, F. F. (1996). *Learning and incentive compatible mechanisms for public provision*. Working paper, University of Michigan.

Cheung, Y. W., & Friedman, D. (1994). *Learning in evolutionary games: Some laboratory results*. Working Paper #303, UC Santa Cruz.

Cheung, Y. W., & Friedman, D. (1996). *A comparison of learning and replicator dynamics using experimental data*. Working Paper #347, UC Santa Cruz.

Edwards, W. (1961). Probability learning in 1000 trials, *Journal of Experimental Psychology*, **62**, 385–394.

Erev, I. (1998). Signal detection by human observers: A cutoff reinforcement learning model of categorization decisions under uncertainty. *Psychological Review*, **105**, 280–298.

Erev, I., Gopher, D., Itkin, R., & Greenshpan, Y. (1995), Toward a generalization of Signal Detection Theory to *n*-person games: The example of two person safety problem. *Journal of Mathematical Psychology*, **39**, 360–375.

Erev, I., & Rapoport, A. (in press). Magic, reinforcement learning and coordination in a market entry game. *Games and Economic Behavior*.

Erev, I., & Roth, A. E. (1996). *On the need for low rationality, cognitive game theory: Reinforcement learning in experimental games with unique, mixed strategy equilibria.* Working paper, University of Pittsburgh.

Erev, I., & Roth, A. E. (in press). Predicting how people play games: Reinforcement learning processes on games with unique mixed strategy equilibrium. *American Economic Review*.

Fudenberg, D., & Levine, D. (1995). *Theory of learning in games.* Manuscript, UCLA Economic Department.

Harley, C. B. (1981). Learning the evolutionary stable strategy. *Journal of Theoretical Biology*, **89**, 611–633.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of Experimental Analysis of Behavior*, **4**, 267–272.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior*, **13**, 244–266.

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

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

Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In Luce, R. D., Bush, R. R., & Galanter, E. (Eds.), *Handbook of Mathematical Psychology*, (Vol. 3), 249–410. New York: Wiley.

March, J. G. (1996). Learning to be risk averse. *Psychological Review*, **103**, 309–319.

McKelvey, R., & Palfrey, T. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, **10**, 6–38.

Mookherjee, D., & Sopher, B. (1997). Learning and decision costs in experimental constant sum games. *Games and Economic Behavior*, **12**, 97–132.

Myers, J. I., Fort, J. G., Katz, L., & Suydam, M. M. (1963). Differential monetary gains and losses and event probability in a two-choice situation. *Journal of Experimental Psychology*, **66**, 521–522.

Premack, D. (1965). Reinforcement theory. In D. Levine (Ed.), *Nebraska symposium on motivation:* 1965. Lincoln: Univ. of Nebraska Press.

Premack, D. (1971). Catching up with common sense or two sides of a generalization: Reinforcement and Punishment. In R. Glaser (Ed.), *The nature of reinforcement*. San Diego: Academic Press.

Rapoport, A., Seale, D., Erev, I., and Sundali, J. A. (1998). Coordination success in market entry games: Tests of equilibrium and adaptive learning models. *Management Science*, **44**, 119–141.

Rapoport, A., Erev, I., Abraham, E. V., & Olson, D. E. (1997). Randomization and adaptive learning in a simplified poker game. *Organizational Behavior and Human Decision Processes*, **69**, 31–49.

Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in intermediate term, *Games and Economic Behavior*, *Special Issue: Nobel Symposium*, **8**, 164–212.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Siegel, S., & Goldstein, D. A. (1959). Decision-making behavior in a two choice uncertain outcome situation. *Journal of Experimental Psychology*, **57**, 37–42.

Siegel, S., Siegel, S., & Andrews, J. M. (1964). *Choice, strategy, and utility*. New York: McGraw-Hill.

Slonim, R., & Roth, A. E. (1997). *Financial incentives and learning in ultimatum and market games: An experiment in the Slovak republic*. Mimeo, University of Pittsburgh.

Tang, F. (1996). *Anticipatory learning in two-person games: An experimental study*. Discussion Paper B-363, University of Bonn.

Thaler, R. (1987). The psychology of choice and assumption of economics. In A. E. Roth (Ed.), *Laboratory experimentation in economics: Six points of view*. Cambridge, UK: Cambridge Univ. Press.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of associative processes in animals, *Psychological Monographs*, **2**.

Tinklepaugh, L. H. (1928). An experimental study of representative factors in monkeys, *Journal of Comparative Psychology*, **8**, 197–236.

Tolman, E. C. (1932). *Purposive behavior in animals and man*. New York: Appleton.

Received: February 4, 1998