

# On Adaptation, Maximization, and Reinforcement Learning Among Cognitive Strategies

Ido Erev

Technion–Israel Institute of Technology

Greg Barron

Harvard Business School

Analysis of binary choice behavior in iterated tasks with immediate feedback reveals robust deviations from maximization that can be described as indications of 3 effects: (a) a *payoff variability effect*, in which high payoff variability seems to move choice behavior toward random choice; (b) *underweighting of rare events*, in which alternatives that yield the best payoffs most of the time are attractive even when they are associated with a lower expected return; and (c) *loss aversion*, in which alternatives that minimize the probability of losses can be more attractive than those that maximize expected payoffs. The results are closer to probability matching than to maximization. Best approximation is provided with a model of reinforcement learning among cognitive strategies (RELACS). This model captures the 3 deviations, the learning curves, and the effect of information on uncertainty avoidance. It outperforms other models in fitting the data and in predicting behavior in other experiments.

*Keywords:* decisions from experience, stickiness effect, case-based reasoning, probability learning, learning in games

Experimental studies of human decision making in iterated tasks reveal a general tendency to respond to immediate feedback in an adaptive fashion. In line with the law of effect (Thorndike, 1898), the probability of successful responses tends to increase with time. Nevertheless, under certain conditions, human adaptation does not ensure maximization. Early research demonstrated robust deviations from maximization that can be summarized with the probability matching assumption (see Estes, 1950, and our discussion below). Under this assumption, the proportion of time an alternative is selected is identical with the proportion of time in which this alternative provides the best outcome.

The main goal of the current research is to improve our understanding of the relationship between adaptation and maximization. In particular, we try to integrate the knowledge accumulated in early studies of probability matching with observations drawn from more recent studies of decisions from experience (e.g., Barron & Erev, 2003; Busemeyer, 1985). We focus on simple situations in which decision makers (DMs) repeatedly face the same binary choice problem and receive immediate feedback after each

choice. The main results of the current analysis are summarized in four sections.

In the first section, we review the known deviations from maximization and show that they can be attributed to three distinct effects. One set of deviations can be classified as indicating a *payoff variability effect* (Busemeyer & Townsend, 1993; Myers & Sadler, 1960): An increase in payoff variability seems to move choice behavior toward random choice. A second set of deviations indicates *underweighting of rare events* (Barron & Erev, 2003): DMs tend to prefer the alternative that provides the best payoff most of the time, even when this alternative is associated with a lower expected return. A third set of deviations involves *loss aversion* (see Kahneman & Tversky, 1979): In certain cases, subjects tend to prefer alternatives that minimize losses over those that maximize payoffs.

The second section highlights the relationship of the results to the probability matching assumption. It shows that this simple assumption provides a good prediction of the main results. Yet the predictions are biased in five ways. The main biases can be captured with an extended probability matching model.

The third section clarifies the implications of the current results to the attempt to develop descriptive learning models. In it, we develop and compare alternative models of the joint effect of the observed behavioral tendencies. The results demonstrate the value of models that assume that the different deviations from maximization can be summarized as negative by-products of three reasonable cognitive strategies. The best fit of the experimental data and of a related thought experiment is provided by a four-parameter model that assumes reinforcement learning occurs among the different cognitive strategies.

In the fourth section, we explore the predictive validity of the proposed models. The results highlight the robustness of the learning models that best fit the data. With the original parameters, these models provide good predictions of the outcomes of 27 tasks studied by Myers, Reilly, and Taub (1961, in a systematic exam-

---

Ido Erev, Max Werthiemi Minerva Center for Cognitive Studies, Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology, Haifa, Israel; Greg Barron, Negotiations, Organizations, and Markets Unit, Harvard Business School.

Part of this research was conducted when Ido Erev was a visiting professor at Columbia Business School. This research was supported by a grant from the National Science Foundation and the USA.–Israel Binational Science Foundation. We thank Ernan Haruvy, Al Roth, Yoav Ganzach, Meira Ben-Gad, and the subjects of seminars at Harvard University, Columbia University, New York University, the University of Michigan, the University of Chicago, and the University of Maryland for useful comments.

Correspondence concerning this article should be addressed to Ido Erev, Max Werthiemi Minerva Center for Cognitive Studies, Faculty of Industrial Engineering and Management, Technion, Haifa, Israel. E-mail: erev@tx.technion.ac.il

ination of the relative effects of incentives and probabilities) and capture the results obtained in studies of repeated plays of matrix games with a unique mixed-strategy equilibrium (the 10 conditions studied in Erev, Roth, Slonim, & Barron, 2002).

### Three Paradigms and the Observed Deviations From Maximization

The current review focuses on three related experimental paradigms. In all three paradigms, the DMs are faced with the same decision problem many times and have limited initial information. The DMs are instructed (and motivated) to try to use the (immediate and unbiased) feedback they receive after each choice to maximize their earnings.<sup>1</sup> The first paradigm considered here, known as *probability learning*, was extensively studied in the 1950s and 1960s in research that focused on the probability matching assumption (see reviews in Estes, 1964, 1976; Lee, 1971; Luce & Suppes, 1965; Shanks, Tunney, & McCarthy, 2002; Vulkan, 2000). In each trial of a typical probability learning study, the DM is asked to predict which of two mutually exclusive events (E or not-E) will occur—for example, which of two lights will be turned on. The probability of each event is static throughout the multitrial experiment. The DMs receive no prior information about probabilities but know the payoff from correct and incorrect predictions. Immediately after each prediction, the DMs can see which event occurred and can calculate their payoffs (and forgone payoffs). The left column in Figure 1 summarizes a typical trial.

The second paradigm involves a choice between two unmarked buttons on the computer screen (see the center column in Figure 1 for an example). In each trial, the DM is asked to click on one of the buttons. Each click leads to a random value drawn from a payoff distribution (a play of a gamble) associated with the selected key. The distributions do not change during the experiment. The DM receives no prior information concerning the relevant payoff distributions but can see the drawn value (the obtained payoff) after each trial. We refer to this basic setting as the *minimal information* paradigm.

The third paradigm is a variant of the minimal information paradigm with more complete feedback. After each choice, the DM is presented with random values drawn from the payoff distributions of each of the two buttons—but payoffs are determined on the basis of the value of the selected button. The additional feedback is often referred to as information concerning forgone payoffs. A typical trial in this *complete feedback* paradigm is presented in the right column of Figure 1.

To demonstrate the observed deviations from maximization,<sup>2</sup> we summarize in the current section the results of 40 experimental conditions, each of which involves at least 200 trials. To facilitate an efficient summary of this large set of data, we focus our analysis on the aggregate proportion of maximization in blocks of 100 trials.

Recall that in the first trial of the experimental conditions considered here, subjects are expected to respond randomly (at an expected maximization rate of around .50). Over the 40 conditions considered here, the rate of maximization in the second experimental block (Pmax2) was over .72. Thus, on average, experience leads toward maximization. Nevertheless, examination of the different conditions reveals that they highlight three classes of deviations from maximization.

### The Payoff Variability Effect

The most obvious class of failures to maximize immediate payoffs involves situations with high payoff variability, such as casino slot machines (see Haruvy, Erev, & Sonsino, 2001). A particularly clear and elegant demonstration of this effect appears in a series of articles by Myers and his associates (see Myers & Sadler, 1960; see also the recent analysis by Busemeyer & Townsend, 1993). A simplified replication of this demonstration is summarized in the top panel of Figure 2. All three problems displayed in this panel present a choice between alternative H (high), with an expected value (EV) of 11 points, and alternative L (low), with an EV of 10 points. The problems differ with respect to the variance of the two payoff distributions:

<i>Problems 1–3 (minimal information, 200 trials, n=14, 0.25¢)</i>			
Problem 1	H	11 points with certainty	Pmax2 = 0.90
	L	10 points with certainty	
Problem 2	H	11 points with certainty	Pmax2 = 0.71
	L	19 points with probability 0.5 1 otherwise	
Problem 3	H	21 points with probability 0.5 1 otherwise	Pmax2 = 0.57
	L	10 points with certainty	

These problems were examined (see Haruvy & Erev, 2001) using the minimal information paradigm in a 200-trial experiment with a conversion rate of 0.25¢ per point. The experimental task was described as the operation of a two-key money machine that was presented on the computer screen. The subjects were told that each selection would lead to an immediate payoff and that their goal was to maximize total earnings. Pmax2 was .90 in Problem 1, .71 in Problem 2, and .57 in Problem 3 (a summary of individual differences and standard deviations is presented below).

Notice that the difference between Problems 1 and 3 appears to reflect risk aversion (H is less attractive when its payoff variability increases), but the difference between Problems 1 and 2 appears to reflect risk-seeking behavior (L is more attractive when its payoff variability increases). This observation suggests that the risk attitude concept found to provide a useful summary of choice behav-

<sup>1</sup> The focus on immediate feedback implies that the current review does not include the important effects of delayed payoff and the related Melioration phenomenon explored by Herrnstein and his associates (see Herrnstein et al., 1993). The relationship between melioration and the results reviewed here is discussed in below under the heading *Boundaries*. In addition, to facilitate the relationship to the economic literature, the current review does not consider studies that did not use monetary incentives (see Hertwig & Ortmann, 2001, for a discussion of this issue).

<sup>2</sup> Notice that in the current settings almost any behavior can be justified as rational; it is possible to find a particular set of prior beliefs that would lead Bayesian agents to exhibit this behavior. One example emerged from informal discussion with the subjects of the experiment. One subject said that he had selected an option that led to bad outcomes in the previous trials because he felt that it was about time for this option to yield good outcomes. Finding the optimal response in the current paradigms is more difficult than in bandit problems (see Berry & Fristedt, 1985) because our DMs are not informed that the payoff distributions are stable. Thus, the current paper focuses on deviations from maximization that may not imply violations of the rationality assumption.









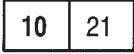
	Experimental Paradigm		
	Probability Learning	Minimal Information	Complete Feedback
Typical Instructions	The current experiment includes many trials. Your task, in each trial, is to guess which of the two light bulbs presented on the screen will light up next. One of the two bulbs will light up immediately after your guess. You will win 4 Agorot if you have guessed correctly, and win nothing otherwise. Your goal is to maximize your total payoff.	The current experiment includes many trials. Your task, in each trial, is to click on one of the two keys presented on the screen. Each click will result in a payoff that will be presented on the selected key, and will be added to your total payoff. Your goal is to maximize your total payoff.	The current experiment includes many trials. Your task, in each trial, is to click on one of the two keys presented on the screen. Each click will result in a payoff that will be presented on the selected key, and will be added to your total payoff. The payoff that you could have obtained from clicking the second key will be presented on that key. Your goal is to maximize your total payoff.
Typical trial			
1. Initial display (the task)	Predict which side will be turned on: 	Click on one of the two keys 	Click on one of the two keys 
2. Response			
3. Display after the response (feedback)	 Right was lit. Payoff for this trial: 0 Total payoff: 2040	 Payoff for this trial: 10 Total payoff: 2040	 Payoff for this trial: 10 Total payoff: 2040

Figure 1. The typical instructions and sequence of displays in a typical trial under the three experimental paradigms.

ior in one-shot tasks, where decisions are made on the basis of a description of the incentive structure, might be less useful in summarizing feedback-based choices. Rather, it seems that the results can be described as a negative side effect of the reasonable tendency to increase exploration in a noisy environment. When the payoff variability is large and exploration is counterproductive, this tendency leads to a payoff variability effect: It moves choice behavior toward random choice. This effect is particularly strong when the variability is associated with the high-EV alternative.

The lower panels of Figure 2 present additional demonstrations and clarifications of the payoff variability effect. The second panel (Problems 4, 5, and 6) shows that the effect is robust to the payoff domain (gain or loss). These problems were run using the same procedure as was used in Problems 1, 2, and 3, with the exception that the DMs received a high show-up fee to ensure similar total expected payoff in the gain (Problems 1–3) and loss (Problems 4–6) domains. The observed robustness implies an important difference between decision making in one-shot, description-based tasks and decision making in

feedback-based decisions. In one-shot tasks, DMs tend to be risk averse in the gain domain and risk seeking in the loss domain. This pattern is referred to as the *reflection effect* (Kahneman & Tversky, 1979). Feedback-based decisions, however, can lead to a reversal of this pattern.

The third panel of Figure 2 (Problems 7–10) shows robustness to the available information. Problems 7–10 are replications of Problems 1, 3, 4, and 5 using the complete feedback paradigm. After each choice, the DMs observed two draws, one from each distribution. The results show that in the current setting, the additional information increases risk seeking. The risky option was more attractive in Problem 8 than in Problem 3; it was also more attractive in Problem 10 than in Problem 5. Yet this effect is rather weak, and it does not change the basic trend. This observation implies that the observed behavior cannot be easily explained as rational exploration in banditlike problems (see Berry & Fristedt, 1985). It is hard to justify persistent exploration in the complete feedback paradigm (Problems 7–10) on the basis of purely rational considerations.

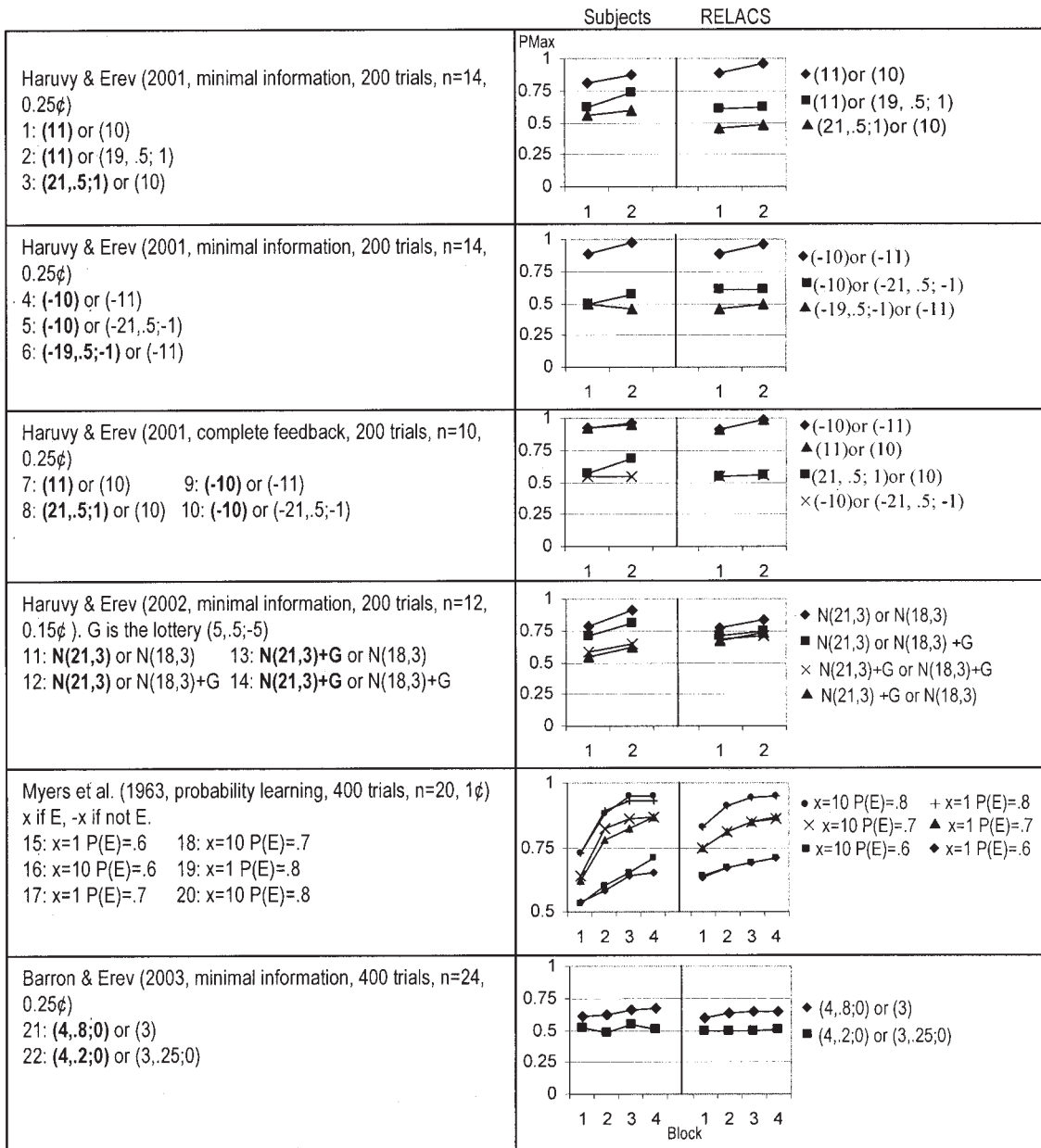


Figure 2. Demonstrations of the payoff variability effect (proportion of maximization [Pmax] in blocks of 100 trials). The notation (x,p;y) describes a gamble that pays x with probability p, y otherwise. N(x,y) means a draw from a normal distribution with mean x and standard deviation y. Boldface highlights the H (high) alternative. RELACS = reinforcement learning among cognitive strategies.

The fourth panel of Figure 2 (Problems 11–14, a replication of Busemeyer, 1985) shows robustness of the effect in a multioutcome environment. These problems, studied by Haruvy and Erev (2002; 200 trials, minimal information,  $n = 12$ , 0.15¢ per point), involve choice among normal distributions. A comparison of Problems 11 and 12 reveals a replication of the observation that an increase in payoff variability can increase the attractiveness of the low-EV gamble.

The fifth panel of Figure 2 (Problems 15–20) summarizes an examination by Myers, Fort, Katz, and Suydam (1963) of the

interaction between the variability effect and payoff magnitude. All six problems used the following format:

Problems 15–20 (probability learning, 400 trials,  $n=20$ ,  $p(E)>0.5$ )  
 H    x if E occurs, -x otherwise  
 L    x if E does not occur, -x otherwise

Myers et al. (1963) manipulated the value of x (1¢ or 10¢) and of P(E) (0.6, 0.7, or 0.8) in a  $2 \times 3$  design. The gambles were presented in a probability learning framework: The DMs were asked to predict which of two mutually exclusive events (E or

not-E) would occur. Correct predictions paid  $x$ , and incorrect responses led to a loss of  $x$  cents. The results show slow adaptive learning and a relatively weak and insignificant payoff magnitude effect.

A payoff variability effect can be observed in one-shot decisions (see Busemeyer & Townsend, 1993). Yet the effect in repeated tasks appears to be more robust. An interesting example is provided by the study of the certainty effect (Kahneman & Tversky, 1979) using a variant of Allais's (1979) common ratio problems. In an earlier article, we (Barron & Erev, 2003) examined the two problems presented in the last panel of Figure 2 (400 trials, minimal information,  $n = 24$ , 0.25¢ per point). Notice that Problem 22 was created by dividing the probability of winning in Problem 21 by 4. This transformation does not affect the prediction of expected utility theory (von Neumann & Morgenstern, 1947) but does affect behavior. In the one-shot task, it reduces the attractiveness of the safer alternative (L). Kahneman and Tversky (1979) called this pattern the *certainty effect*. It is interesting that in the repeated task, the division by 4 increases the attractiveness of the safer alternative. We noted that the repeated choice pattern can be a result of the payoff variability effect: The payoff variability (relative to the differences in expected value) is larger in Problem 22 than in Problem 21. As a result, DMs are less sensitive to the expected values and behave as if L is more attractive.

*Underweighting of Rare Outcomes*

A second class of problems in which adaptation does not lead to maximization involves situations in which the alternative that provides the best outcome most of the time has the lower expected payoff. In these situations, the reasonable tendency to rely on the typical outcomes implies underweighting of rare (low probability) outcomes. One indication of this pattern is provided by the study of Problem 23 using the minimal information paradigm (in Barron & Erev, 2003).

Notice that H has a higher expected value (3.2 vs. 3), but in most (90%) of the trials, L provides better payoff (3 vs. 0). The proportion of maximization in Trials 101 to 200 was only 0.24. Moreover, initially, the maximization rate decreased with experience (see Figure 3).

Figure 3 shows two additional conditions (run using the procedure of Problem 23) that demonstrate the robustness of the tendency to underweight rare events. Problem 24 shows this pattern

when both alternatives are risky. Problem 25 shows a similar pattern in the loss domain.

Recall that studies of decision making in one-shot tasks reveal a strong tendency to overweight small probabilities, leading to important deviations from maximization. This tendency is captured by the weighting function of prospect theory (see Gonzalez & Wu, 1999, for a careful analysis of this concept). The results of all three problems summarized in Figure 3 show that in feedback-based directions, DMs tend to deviate from maximization in the opposite direction. On average, DMs behave as if they underweight rare events.

The results summarized in Figure 3 could also be captured with the assumption of an extreme reflection effect (risk aversion in the gain domain and risk seeking in the loss domain; see Kahneman & Tversky, 1979). Yet this assumption is inconsistent with the results summarized in Figure 2. In addition, Barron and Erev (2003) observed a preference for the possibility "loss of 9 with certainty" over the gamble "loss of 10 with probability 0.9, 0 otherwise." This pattern contradicts the prediction of the reflection hypothesis and is predicted by the assumption that low-probability outcomes (the 0.1 probability to avoid losses) are underweighted.

Examination of how the availability of information concerning forgone payoffs affects the tendency to underweight rare events supports the suggestion that this information increases risk seeking. Information concerning forgone payoffs was found to reduce sensitivity to unattractive rare events (see Yechiam & Busemeyer, 2005) while increasing sensitivity to attractive rare events (see Grosskopf, Erev, & Yechiam, 2005). Yet both effects are relatively mild. The tendency to underweight rare events is robust over the different experimental paradigms.

Barron and Erev (2003) noted that the underweighting of rare events in repeated decisions can be a product of a tendency to rely on recent experiences and/or reliance on small samples (see Fiedler, 2000; Kareev, 2000). According to the recency argument, rare events are underweighted because they are not likely to have occurred recently. Examination of this hypothesis shows that the recency effect is robust but is only one contributor to the tendency to underweight rare events. For example, in Problem 25 ("loss of 3 with certainty" or "loss of 32 with probability of 0.1; 0 otherwise"), 21 of the 24 subjects exhibited the recency effect: They were less likely to select the gamble again after a loss of 32. Yet most subjects (14 of 24) selected the gamble again (in more than

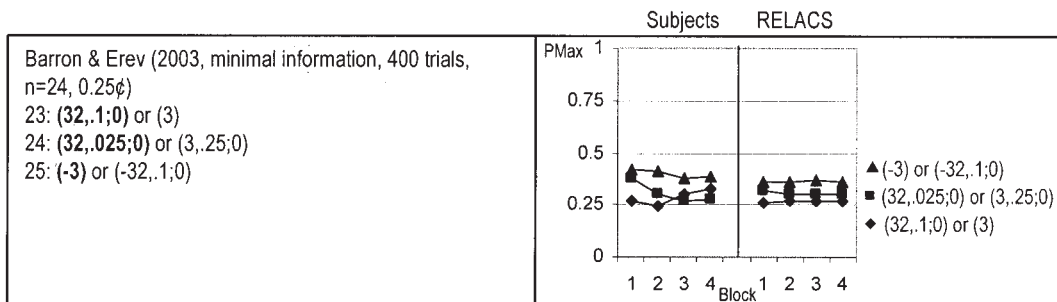


Figure 3. Demonstrations of underweighting of rare events (proportion of maximization [Pmax] in blocks of 100 trials). Boldface highlights the H (high) alternative. RELACS = reinforcement learning among cognitive strategies.

50% of the cases) even immediately after a loss of 32. This observation can be described with the suggestion that at least part of insensitivity to rare events is a result of reliance on small (and not necessarily recent) set of experiences.

### *Loss Aversion and Probability Matching*

Studies of behavior in the stock market have led to the discovery of a third class of maximization failures. Thaler, Tversky, Kahneman, and Schwartz (1997, and see Gneezy & Potters, 1997) showed that when the action that maximizes expected value increases the probability of losses, people tend to avoid it. That is, the reasonable tendency to avoid losses can lead to a counterproductive loss aversion. This idea helps explain the observation that many investors are underinvested in the stock market (relative to the prescription of reasonable models). In a replication of these studies, Barron and Erev (2003) considered the following problem:

- Problem 26 (minimal information, 200 trials,  $n=12$ , 1¢ per 100 points)*
- H A draw from a normal distribution with a mean of 100 and standard deviation of 354.  $P_{max2} = 0.32$
- L A draw from a truncated (at 0) normal distribution with a mean of 25 and standard deviation of 17.7. (Implied mean of 25.63.)

The proportion of maximization after 200 trials with immediate payoff was 0.32. Similar results were obtained in the original studies (Gneezy & Potters, 1997; Thaler et al., 1997) in which the subjects received more information (including forgone payoffs).

To establish that the suboptimal behavior observed in Problem 26 is a result of loss aversion, Thaler et al. (1997) compared this condition with an inflated condition. In Barron and Erev's (2003) replication, adding 1,200 to the means of the two distributions created the inflated condition. That is,

- Problem 27 (minimal information, 200 trials,  $n=12$ , 1¢ per 100 points)*
- H A draw from a normal distribution with a mean of 1,300 and standard deviation of 354.  $P_{max2} = 0.56$
- L A draw from a normal distribution with a mean of 1,225 and standard deviation of 17.7.

The elimination of the losses increased the proportion of maximization in the last block to 0.56. The top panel in Figure 4 summarizes the learning curves in the two replications of Thaler et al. described above. In addition, Figure 4 shows a third condition (Problem 28) that reveals a further and larger increase in maximization, obtained by reducing payoff variability.

Thaler et al. (1997) showed that their results are consistent with the predictions of a myopic version of prospect theory (Kahneman & Tversky, 1979). According to this theory, loss aversion is a reflection of the fact that the subjective enjoyment from gaining a certain amount tends to be smaller than the subjective pain from losing the same amount. Under this interpretation of Thaler et al.'s results, loss aversion in repeated-choice tasks resembles loss aversion in one-shot decisions from experience (the decisions captured by prospect theory). However, other studies of repeated decisions show important differences between the tendency to avoid losses in one-shot decisions and loss avoidance in repeated-choice tasks. A clear demonstration of this difference is provided in Katz (1964). He found, for example, that in repeated tasks, decision makers are indifferent between "equal chance for 1 and -1" and

"equal chance for 4 and -4." Thus, they behave as if their value function has the same slope in the gain and the loss domains.

Additional data concerning the effect of losses come from probability learning studies that manipulated payoff signs. Seven of the experimental conditions considered here focused on the following task:

- Problems 29–35 (probability learning, 400–500 trials,  $P(E)>0.5$ )*
- H G if E occurs, B otherwise
- L G if E does not occur, B otherwise

The second panel in Figure 4 summarizes Siegel and Goldstein's (1959) study that compared two problems with  $P(E) = 0.75$  and  $G = 5¢$ . The value of B was 0 in Problem 29 and  $-5¢$  in Problem 30. Thus, G stands for the good payoff, and B stands for the (relatively) bad payoff. The proportion of maximization in the second block was 0.85 and 0.95 in Problems 29 and 30, respectively.

The higher maximization rate in Problem 30 can stem from the larger difference between G and B (5 or 10) and/or a positive effect of the losses. To evaluate the possibility that losses can facilitate maximization, Erev, Bereby-Meyer, and Roth (1999, and see Bereby-Meyer & Erev, 1998) examined Problems 31–35 ( $n = 14$ ). They used the parameters  $P(E) = 0.7$ ;  $G = 6, 4, 2, 0$  or  $-2$ ; and  $B = G - 4$  (with a conversion rate of 0.25¢ per point). The results, presented in the third panel of Figure 4, show a nonlinear effect of G on learning speed: Maximal learning speed was observed when G was 2 or 0.

To compare alternative explanations of the nonlinear effect of G in Problems 30–34, Erev et al. (1999) ran four additional conditions that can be summarized as follows:

- Problems 36–39 (probability learning and minimal information,  $n=9$ , 550 trials,  $P(E)=0.7$ ,  $P(F)=0.9$ , 0.25¢ per point)*
- H G if E and F occur, B if not-E and F occur, 0 otherwise
- L G if not-E and F occurs, B if E and F occur, 0 otherwise

The value of G was 6 (Problems 36 and 38) or  $-2$  (Problems 37 and 39). As in Problems 31–35,  $B = G - 4$ . Problems 36 and 37 were studied in a probability-learning paradigm and Problems 38 and 39 in a minimal information paradigm. The results (displayed in the fourth panel of Figure 4) show very small (and insignificant) differences between the four conditions.

An evaluation of Problems 26–39 suggests that the main results can be summarized as a loss-aversion-probability matching effect: That is, the availability of an alternative that minimizes the probability of losses moves choice behavior from probability matching toward preferring this alternative. Recall that probability matching implies a matching of the choice probabilities to the proportion of trials in which the selected alternative maximizes earnings. In Problems 30, 33, and 34, probability matching implies a lower maximization rate (75% or 70%) than does selecting the action that minimizes the probability of losses (Action H). In Problem 27, however, probability matching implies a 58% maximization rate, and the action that minimizes the probability of losses is L.

The lower panel in Figure 4 presents a comparison that highlights the possible trade-offs and interactions between loss aversion and the payoff variability effect. Notice that the payoff variability effect implies a higher maximization rate in Problem 40 than in Problem 21 (because in Problem 40, H involves less variability). Loss aversion implies the opposite (because in Problem 40, H involves a higher loss rate). The results (minimal

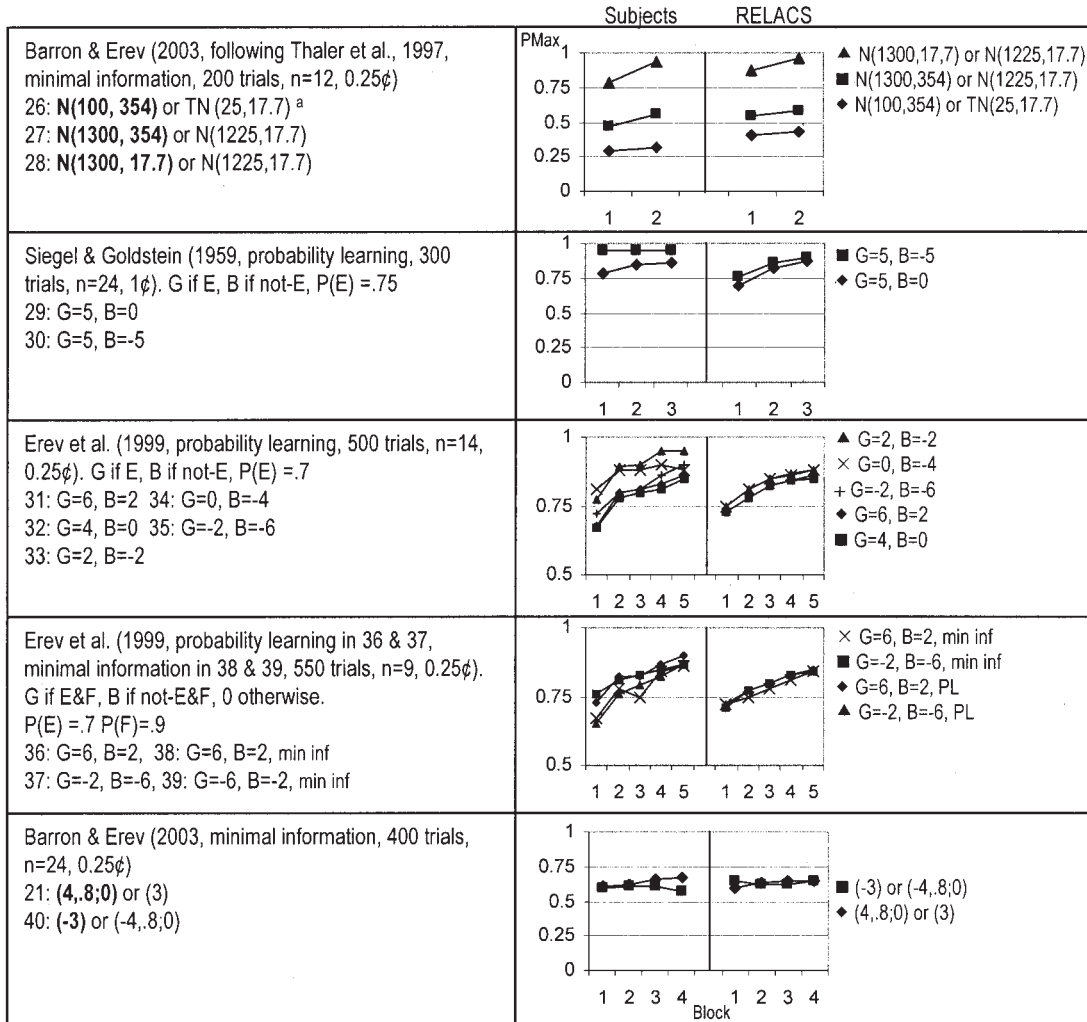


Figure 4. Demonstrations of the loss rate effect (proportion of maximization [Pmax] in blocks of 100 trials). N(x, y) means a draw from a normal distribution with mean x and standard deviation y. E and not-E are two mutually exclusive events. Boldface highlights the H (high) alternative. <sup>a</sup>TN(25, 17.7) is a truncated (at 0) normal distribution.

information, 400 trials,  $n = 24$ ,  $0.25\phi$  per point) show an insignificant difference between the two problems. Thus, in the current example, the two effects appear to cancel each other out.

Comparison of the Three Paradigms

The previous section focuses on behavioral regularities (deviations from maximization) that appear to occur under all three experimental paradigms considered here. Obviously, however, there are also interesting differences between the paradigms. Two differences are particularly relevant to the current analysis. The first of these involves a correlation effect. As demonstrated by Diederich and Busemeyer (1999), when the payoffs of the two alternatives are positively correlated, the availability of information concerning forgone payoffs eliminates the payoff variability effect. Specifically, when Alternative H dominates L in all trials, payoff variability has little effect.

The second difference involves the effect of forgone payoffs on implied risk attitude. As noted above, the availability of informa-

tion about forgone payoffs increases the tendency to select the riskier alternative. This pattern can be described as the outcome of a stickiness effect: In the absence of information about forgone payoffs, sequences of bad outcomes (that are likely to come from the riskier alternative) have a lasting effect because they inhibit future updating of the tendency to select this alternative. This effect is an example of the more general hot stove effect highlighted in Denrell and March (2001).

The Relationship to Probability Matching

Early attempts to summarize the behavioral regularities observed in experimental studies of iterated decision tasks have demonstrated the value of the probability matching assumption (see Estes, 1950). To clarify the relationship of the current results to probability matching, we derived the predictions of this assumption to the experimental conditions summarized above. The predictions were derived by computing (for each problem) the probability that Alternative H provides the best outcome in one

randomly selected trial. According to the probability-matching assumption, the observed H rate matches this probability. Notice that the probability-matching assumption implies sensitivity to forgone payoffs and to the correlation between the outcomes of the two alternatives. When forgone payoffs are known, the best alternative is determined by comparing the outcomes in the same trial. When forgone payoffs are unknown, the computation of the best alternative is based on pairing different random trials for each alternative.

Table 1 presents the 40 experimental conditions considered above (in the leftmost column), the observed H rate in the second block (in the Pmax2 column), and (in the Probability matching column) the predicted H rate under the probability matching assumption. The results show high correlation ( $r = .81$ ) between the observed and predicted proportions (see the related observation in Blavatsky, 2004). The mean squared deviation (MSD) between the observed and predicted rates is 0.017. This value is much better than the MSD of the maximization assumption (0.120) and the expected MSD under random choice (0.080).

In addition, the results reveal five types of deviations from probability matching. First, in most cases (27 out of 40), the observed maximization rate is larger than the probability matching predictions (see early observations in Edwards, 1961; Siegel & Goldstein, 1959). A second bias involves situations in which probability matching implies maximization. In these cases (Problems 1, 4, 7, 9, and 28), the observed maximization rate deviates from this prediction in the direction of random choice. A third bias emerges in the minimal information paradigm when H has higher payoff variability than L and probability matching implies a moderate (38% to 80%) maximization rate. In six of the seven problems with these characteristics (Problems 6, 21, 22, 24, 26, and 27; the sole exception is Problem 3), the observed maximization rate is lower than the probability matching predictions. This bias can be described as an indication of the stickiness effect. A fourth bias is suggested by Problem 40 and the larger deviation from probability matching in Problem 26 relative to Problem 27: a deviation from probability matching toward the low-EV alternative that minimizes the probability of losses. Finally, Problem 13 reflects a small deviation toward random choice that can be described as an indication of slow learning in situations with high payoff variability.

#### *An Extended Probability Matching Model*

The observation that the predictions of the probability matching assumption are useful (highly correlated with the data) but biased in easy-to-describe ways suggests that it should be easy to refine this assumption so as to capture the data without losing the original insight. To explore this possibility, we considered (and present in Table 1) six variants of the probability-matching assumption. The first variant, referred to as PM2, assumes that the observed H rate matches the proportion of time in which Alternative H provides the best outcome in a sample of two randomly selected trials.

We also consider extensions of this idea with larger sample sizes. Variant PM $k$  assumes that the observed H rate matches the proportion of time in which Alternative H provides the best outcome in a sample of  $k$  randomly selected trials.

The final variant of probability matching considered here was motivated by the data summarized in Figure 4. It assumes there is an interaction between loss aversion and probability matching.

Under this variant, referred to as PM-LA, the DMs follow the probability matching predictions when the two alternatives lead to identical loss rates but prefer the alternative with the lower loss rate otherwise.

#### *High Correlation and Five Deviations*

The bottom panel in Table 1 presents the correlation between the different variants and the observed H rate, as well as the relevant MSD scores. The results show that the data are best captured with the PM3 and PM4 predictions. This observation is consistent with recent demonstrations of the tendency to rely on small samples (see Fiedler, 2000; Kareev, 2000). When all seven variants are entered into a regression analysis, five of the seven can be eliminated without impairing the fit. The predictors that cannot be eliminated are PM4 and PM-LA ( $p < .05$ ). The advantage of these predictors is robust to the method of analysis (linear regression or logistic regression).

To simplify the interpretation of the regression analysis, we focus on a linear model that allows random choice and that predicts rates between 0 and 1. That is, we focus on models of the type Prediction =  $b_1(\text{PM4}) + b_2(\text{PM-LA}) + b_3(0.5)$  with the constraint  $b_1 + b_2 + b_3 = 1$ . The estimated free parameters are  $b_1 = .62$  and  $b_2 = .22$  (and the hypothesis  $b_1 + b_2 + b_3 = 1$  cannot be rejected). Thus, our favorite variant of the probability-matching assumption, referred to as the *extended probability matching* (EPM) model, assumes that the proportion of H choices is given by  $P(H) = .62(\text{PM4}) + .22(\text{PM-LA}) + .16(0.5)$ . The correlation of this model with the data (Pmax2) is .94, and the MSD is 0.005.

#### *An Abstraction of the Learning Process*

The main shortcoming of the EPM model presented above is its static nature. This model cannot capture either the stickiness effect or the observed increase in maximization with experience. To capture the observed learning curves, we explore in the current section models that describe the basic ideas of the EPM model as properties of a learning process.

#### *Reinforcement Learning Among Cognitive Strategies (RELACS)*

The first learning model considered here assumes that reinforcement learning occurs among cognitive strategies (see Erev & Roth, 1999, and related ideas in Busemeyer & Myung, 1992; Erev, 1994, 1998; Luce, 1959; Payne, Bettman, & Johnson, 1993; Skinner, 1953; Stahl, 1996). Specifically, the model assumes that in each trial, the DM follows one of three cognitive strategies (decision rules). The probability that any given rule will be used is determined by reinforcements derived from previous experiences with the rule. The model, referred to as RELACS, can be summarized with the following assumptions.

#### *Fast Best Reply*

The observation that the PM4 rule provides good fit to the current data is captured here by the assumption that one of the cognitive strategies considered by the decision makers is a weighted adjustment rule (and see similar abstractions in Bush & Mosteller, 1955; Estes, 1950; March, 1996; Sarin & Vahid, 2001):

Table 1  
Evaluation of the Probability Matching Predictions

Problem and paradigm	Alternative H	Alternative L	Pmax2	Probability matching	Variants of probability matching (PM)						
					PM2	PM3	PM4	PM5	PM6	PM-LA	EPM
1 MI	(11)	(10)	0.87*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92
2 MI	(11)	(19, .5; 1)	0.74	0.50	0.75	0.50	0.69	0.50	0.66	0.50	0.62
3 MI	(21, .5; 1)	(10)	0.59	0.50	0.75	0.50	0.69	0.50	0.66	0.50	0.62
4 MI	(-10)	(-11)	0.97*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92
5 MI	(-10)	(21, .5; -1)	0.57	0.50	0.75	0.50	0.69	0.50	0.65	0.50	0.62
6 MI	(-19, .5; -1)	(-11)	0.46*	0.50	0.75	0.50	0.69	0.50	0.65	0.50	0.62
7 CF	(11)	(10)	0.96*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92
8 CF	(21, .5; 1)	10	0.68	0.50	0.75	0.50	0.69	0.50	0.66	0.50	0.62
9 CF	-10	-11	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92
10 CF	-10	(-21, .5; -1)	0.55	0.50	0.75	0.50	0.68	0.50	0.66	0.50	0.61
Problems 11-14: G is the gamble (-5, .5; +5)											
11 MI	N(21,3)	N(18,3)	0.91	0.76	0.84	0.89	0.92	0.94	0.96	0.76	0.82
12 MI	N(21,3)	N(18,3) + G	0.81	0.64	0.73	0.78	0.81	0.85	0.87	0.64	0.73
13 MI	N(21,3) + G	N(18,3)	0.62*	0.64	0.73	0.78	0.81	0.84	0.87	0.64	0.73
14 MI	N(21,3) + G	N(18,3) + G	0.65	0.64	0.69	0.73	0.76	0.79	0.81	1.00	0.77
Problems 15-20: H = (x if E; -x if not-E); L = (-x if E; x if not-E)											
15 PL	x = 1, P(E) = .6		0.58*	0.60	0.65	0.68	0.71	0.73	0.75	1.00	0.74
16 PL	x = 10, P(E) = .6		0.60	0.60	0.65	0.68	0.71	0.74	0.75	1.00	0.74
17 PL	x = 1, P(E) = .7		0.78	0.70	0.79	0.84	0.88	0.90	0.92	1.00	0.84
18 PL	x = 10, P(E) = .7		0.82	0.70	0.78	0.84	0.87	0.90	0.92	1.00	0.84
19 PL	x = 1, P(E) = .8		0.89	0.80	0.89	0.94	0.97	0.98	0.99	1.00	0.90
20 PL	x = 10, P(E) = .8		0.88	0.80	0.89	0.94	0.97	0.98	0.99	1.00	0.90
21 MI	(4, 0.80; 0)	(3)	0.62*	0.80	0.64	0.52	0.62	0.74	0.65	0.80	0.64
22 MI	(4, 0.20; 0)	(3, 0.25; 0)	0.48*	0.50	0.52	0.53	0.54	0.55	0.54	0.50	0.52
23 MI	(32; 0.10; 0)	(3)	0.24	0.10	0.19	0.28	0.35	0.41	0.47	0.10	0.32
24 MI	(32, 0.025; 0)	(3, 0.25; 0)	0.30*	0.39	0.32	0.27	0.24	0.22	0.22	0.39	0.32
25 MI	(-3)	(-32, 0.10; 0)	0.41	0.10	0.19	0.27	0.34	0.41	0.47	0.10	0.32
26 MI	N(100, 354)	TN(25, 17.7)	0.32*	0.58	0.62	0.64	0.66	0.68	0.70	0.00	0.49
27 MI	N(1300, 354)	N(1225, 17.7)	0.56*	0.58	0.62	0.64	0.66	0.68	0.70	0.58	0.62
28 MI	N(1300, 17.7)	N(1225, 17.7)	0.94*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92
Problems 29-35: H = (G if E; B if not-E); L = (B if E; G if not-E)											
29 PL	P(E) = .75, G = 5, B = 0		0.85	0.75	0.84	0.90	0.93	0.95	0.97	0.75	0.82
30 PL	P(E) = .75, G = 5, B = -5		0.95	0.75	0.84	0.89	0.93	0.95	0.97	1.00	0.88
31 PL	P(E) = .7, G = 6, B = 2		0.80	0.70	0.78	0.84	0.87	0.90	0.92	0.70	0.78
32 PL	P(E) = .7, G = 4, B = 0		0.78	0.70	0.78	0.84	0.87	0.90	0.92	0.70	0.78
33 PL	P(E) = .7, G = 2, B = -2		0.89	0.70	0.78	0.84	0.87	0.90	0.92	1.00	0.84
34 PL	P(E) = .7, G = 0, B = -4		0.88	0.70	0.78	0.84	0.87	0.90	0.92	1.00	0.84
35 PL	P(E) = .7, G = -2, B = -6		0.80	0.70	0.78	0.84	0.87	0.90	0.92	0.70	0.77
Problems 36-39: P(E) = .7; P(F) = .9; H = (G if E&F, B if not-E&F, 0 otherwise); L = (B if E&F, G if not-E&F, 0 otherwise)											
36 PL	G = 6, B = 2		0.82	0.68	0.76	0.82	0.86	0.89	0.91	0.68	0.76
37 PL	G = -2, B = -6		0.81	0.68	0.77	0.82	0.86	0.89	0.91	0.68	0.76
38 MI	G = 6, B = 2		0.76	0.66	0.74	0.78	0.82	0.85	0.87	0.66	0.73
39 MI	G = -2, B = -6		0.78	0.66	0.74	0.79	0.82	0.85	0.87	0.66	0.81
40 MI	(-3)	(-4, 0.80; 0)	0.61*	0.80	0.64	0.51	0.61	0.74	0.65	0.80	0.64
	Correlation with Pmax2			0.81	0.86	0.89	0.92	0.87	0.90	0.81	0.94
	MSD (×100)			1.71	1.07	0.92	1.03	1.42	1.56	2.70	0.50

Note. The left-hand columns present the 40 problems and the observed maximization rate in the second block (Pmax2). MI = minimal information; CF = complete feedback; PL = probability learning; E and not-E = two mutually exclusive events; B = the relatively bad payoff; MSD = mean squared deviation; PM-LA = probability matching with an interaction with loss aversion; H = high; L = low. The notation (x,p;y) describes a gamble that pays x with probability p, y otherwise. The notation (x if E; y if not-E) implies a gamble that pays x if E occurs and y otherwise. N(x,y) means a draw from a normal distribution with mean x and standard deviation y, TN(x, y) is a truncated (at zero) normal distribution. Asterisks (\*) stand for lower maximization rate than the prediction of probability matching. The right-hand columns present the prediction of the probability matching assumption, different variations of this assumption, and the extended probability matching model (EPM).

*Assumption 1.* In certain trials, the DM follows a fast best reply strategy that implies the selection of the action with the highest recent payoff. The recent payoff of action  $j$  is:

$$R_j(t + 1) = R_j(t)[1 - \beta] + v(t)_j\beta \tag{1}$$

where  $v(t)$  is the observed payoff from  $j$  in trial  $t$ , and  $\beta$  ( $0 < \beta < 1$ ) is a recency parameter: High values imply extreme overweighting of recent outcomes. Random choice is assumed when both actions have the same recent payoff. The recent value of action  $j$  is not updated in trials in which the payoff from  $j$  is not observed. The initial value  $R_j(1)$  is assumed to equal the expected payoff from random choice (e.g., 10.5 in Problem 2).

When forgone payoffs are available, the correlation between the predictions of this rule (with  $\beta = .5$ ) and the prediction of the PM4 rule is .98. Thus, the two rules capture very similar tendencies. The fast best reply abstraction is preferred here because it is consistent with the evidence of recency and stickiness effects considered above.

*Loss Aversion and Case-Based Reasoning*

The relationship between payoff sign and probability matching (and the value of the PM-LA rule) is assumed here to result from the use of a case-based decision rule. This rule implies an initial tendency to select whatever action led to the best outcome in a similar case in the past (see related ideas in Gilboa & Schmeidler, 1995; Logan, 1988; Riesbeck & Schank, 1989). Because all previous trials (cases) are objectively equally similar (in the conditions considered here), the current implementation of this rule assumes that one of the similar cases is randomly selected.

Loss aversion is abstracted with the assertion that when the initial tendency implies a selection of the action associated with higher recalled loss rate, it is reversed.

*Assumption 2.* One of the strategies considered by the DM involves case-based reasoning with a loss-minimizing consistency check. When this strategy is used before observing at least one outcome from each action (or if all previous outcomes were identical), it implies random choice. In other situations, this strategy implies selection of the action expected to be more attractive on the basis of the following two-stage belief-forming rule.

In the first stage, one of the previous trials is (randomly) recalled, and the observed outcomes in that trial form the beliefs for the current trial. If the forgone payoffs from that trial are not known, a second trial is (randomly) recalled from the set of trials for which the DM has knowledge of payoffs from the action not observed at the first recalled trial. If the two actions have the same payoffs in the sample (but are not identical to all previous payoffs), the operation is repeated to break the tie.

In the second stage, the DM is assumed to check the beliefs formed in the first stage and reject them if the action with the higher payoff in the initial stage is recalled to be associated with more frequent and larger average losses. To perform this “loss consistency test,” the DM recalls  $\kappa$  additional past outcomes and computes the number of losses and the total losses from each action over the  $\kappa + 1$  recalled trials per action. The exact number,  $\kappa = 0, 1, 2, \dots$ , is a free parameter that captures sensitivity to rare losses. If both measures point in the opposite direction of the initial belief (i.e., if the action with the higher payoff in the first stage is associated with more losses and higher total losses over the  $\kappa + 1$

trials), the belief is reversed. In other words, the action associated with fewer and smaller losses is believed to be more attractive.

Notice that this case-based rule is a refined version of the PM-LA rule. The refinement involves the assertion that the LA criteria are less likely to be used when it is not easy to see which alternative minimizes losses. In the current setting, the correlation between the two rules (with  $\kappa = 10$ ) is .98.

*Diminishing Random Choice and Slow Best Reply With Exploration*

The observations that a certain percentage of random choices improves the fit of the EPM model and that maximization tends to increase with experience are captured here with the assumption that one of the strategies considered by the DMs involves a slow best reply process. This strategy implies there is approximately random choice at the beginning of the learning process and a slow learning toward preferring the strategy likely to maximize earnings. The learning speed is assumed to depend on the payoff variability effect (see similar abstractions in Busemeyer, 1985; Erev et al., 1999; Friedman & Mezzetti, 2001; Myers & Sadler, 1960; Weber, Shafir, & Blais, 2004).

*Assumption 3.* The third strategy considered by the DM can be abstracted as a slow best reply rule. This strategy assumes a stochastic response rule that implies continuous but diminishing exploration. The probability that alternative (action)  $j$  is taken at trial  $t$  is

$$p_j(t) = e^{W_j(t)\lambda/S(t)} \bigg/ \sum_{k=1}^2 (e^{W_k(t)\lambda/S(t)}) \tag{2}$$

where  $\lambda$  is an exploitation–exploration parameter (low values imply more exploration),  $W_j(t)$  is the weighted average payoff associated with alternative  $j$ , and  $S(t)$  is a measure of payoff variability.

The weighted average rewards are computed like the recent payoffs with the exception of a slower updating parameter  $\alpha$  ( $0 < \alpha < \beta$ ). That is,  $W_j(1) = R_j(1)$  and

$$W_j(t + 1) = W_j(t)[1 - \alpha] + v(t)_j\alpha \tag{3}$$

The initial value of the payoff variability term,  $S(1)$ , is computed as the expected absolute difference between the obtained and expected payoffs from random choice. For example, in Problem 2, where  $H = 11$  and  $L = 19$  with a probability of 0.5, 1 otherwise,  $S(1) = 0.5\text{Abs}(11 - 10.5) + 0.5[0.5\text{Abs}(1 - 10.5) + 0.5\text{Abs}(19 - 10.5)] = 4.75$ . This definition implies a large initial payoff variability effect. To capture the observation that the payoff variability effect is more sensitive to the variability of Alternative H, we assume that the payoff variability measure moves toward the observed mean absolute difference between the obtained payoff  $v(t)$  and the maximum of the last observed payoffs from the two actions ( $\text{Last}_1$  and  $\text{Last}_2$ ). That is,

$$S(t + 1) = S(t)[1 - \alpha] + \text{ABS}[v(t) - \text{Max}(\text{Last}_1, \text{Last}_2)]\alpha \tag{4}$$

Notice that when forgone payoffs are known and the selected action has a higher payoff,  $v(t) = \text{Max}(\text{Last}_1, \text{Last}_2)$ . Thus, in this case,  $S(t + 1)$  moves toward 0. In addition, the current

abstraction captures the effect of correlated outcomes. When forgone payoffs are known, positive correlation reduces the expected value of  $S(t + 1)$ .

### *Choice Among Strategies*

The final assumption implies reinforcement learning among the cognitive strategies.

*Assumption 4.* Choice among strategies follows the stochastic choice rule described in Assumption 3, with one exception: The strategy's weighted average is updated only in trials in which the strategy was used. As in Assumption 3, the initial values are assumed to equal the expected payoff from random choice.

In other words, when a strategy leads to a desired outcome (the outcome is higher than the current propensity), the probability that it is used again increases. An undesired outcome has the opposite effect. This updating process is assumed to occur only after a selection of the strategy because the derivation of the forgone payoff from an unselected strategy requires costly computations. The DM has first to derive the implications of each of the three strategies that are updated according to the observed payoffs. Thus, it is natural to assume that these computations are not made.

### *Descriptive Value*

To fit the model (i.e., estimate the parameters), we ran computer simulations in which virtual DMs that behave according to the four assumptions discussed above participated in a virtual replication of the 40 experimental conditions. (To reduce the risk of programming errors, the simulations were run using two independent computer programs, one in SAS [Version 9.1] and the other in Visual Basic [Version 6]). The four parameters were fitted by considering a wide set of possible values (on the space  $0 < \alpha < 1$ ,  $0 < \lambda < 20$ ,  $\alpha < \beta < 1$ , and  $1 < \kappa < 20$ ). Then 200 simulations were run in each of the 40 problems under each set of parameter values. The following steps were taken in trial  $t$  of each simulation:

1. The virtual DM selected a cognitive strategy that determined an action.
2. The payoff of the observed actions was determined on the basis of the problem payoff rule. (To facilitate comparison with models that imply sensitivity to payoff magnitude, we converted the payoffs to their current value in cents. The conversion rate from the late 1950s–early 1960s data was 1:5.)
3. The maximization rate was recorded.
4. The values assumed to affect future choices were updated.

Each set of simulations resulted in a predicted learning curve in each of the 40 problems under each set of parameters. Like the experimental data set, the predictions were summarized by the proportions of Pmax in blocks of 100 trials. The fit of each set of simulations was summarized by the MSD between the aggregated observed and the predicted curves. The average MSD of each problem (over the 2 to 5 blocks) received the same weight. The parameters that minimize this score are  $\lambda = 8$ ,  $\alpha = .00125$ ,  $\beta = 0.2$ , and  $\kappa = 4$ .

Table 2 summarizes the MSD scores of the models studied here (the first MSD column shows the results for the current data set, while the other columns and the lower rows show generalization scores and scores of alternative models, described below). It shows that the score of RELACS is 0.0036. This score implies that the average distance between the observed and predicted proportions (that can assume values between 0 to 1) is below .06. A more informative evaluation of the magnitude of this score is provided below.

The cells on the right side of Figures 2–5 show the learning curve implied by the model. A comparison of the data with the model reveals good qualitative fit. For example, the model reproduces the nonmonotonic trend observed in Figure 3 and the top panel of Figure 4. In Problems 23, 25, and 26, the model captures the initial learning to prefer L, as well as the slight increase in the proportion of H choices with experience.

In an initial evaluation of the model, it was compared with four baseline models: the EPM rule described above, expected value maximization (Pmax of 1), random choice (Pmax of 0.50), and a five-parameter model that fits the mean of each block (over the 40 curves) to each of the curves. As noted above, the MSD of EPM for the second block (the block used to fit this model) is 0.005, and the MSD of RELACS in that block is 0.0035. This comparison suggests that the added assumptions needed to capture the dynamic do not impair the fit. The MSD scores of the other baseline models (cf. second panel in Table 1) are 0.0792, 0.0528, and 0.0211, respectively. The advantage of RELACS over the five-parameter baseline suggests that it captures some of the robust differences between the different problems.

### *Simpler Models and the Role of the Different Assumptions*

Three sets of analyses were conducted in an attempt to clarify the relative importance of the different assumptions captured in RELACS. The first set of analyses, summarized in the second section of Table 2, focuses on simpler models proposed in previous research. Those models, which include stochastic fictitious play (see Cheung & Friedman, 1998; Fudenberg & Levine, 1998) and reinforcement learning (see Roth & Erev, 1995), can be described as variants of a simplification of RELACS that assume that only one cognitive strategy, slow best reply, is used in all trials. The results of these analyses highlight the importance of the abstraction of the payoff variability effect, the division by  $S(t)$ . Table 2 shows that this assumption reduces the MSD of the simple models by more than 50%. In addition, the results suggest that these simple models are too simple. Even with abstraction of the payoff variability effect, their MSD scores are higher than the RELACS score by more than 100%. The failures of the basic models suggest that they tend to overpredict the stickiness effect.

A second set of analyses focuses on versions of RELACS that assume that only two of the cognitive strategies are used. Comparison of the different models shows, again, the importance of the slow best reply rule that captures the payoff variability effect. The abstraction of fast best reply seems to be less important: The MSD of the model that ignores this tendency is 0.0053.

The third set of analyses considers variants of RELACS that abstract all three cognitive strategies. These analyses lead to the surprising finding that the assumption of learning among the different strategies is not very important in the current context. That is, a model that assumes random choice among the three cognitive strategies fits the data almost as well as RELACS does. Another interesting observation is the weak sensitivity to the

Table 2  
Summary of Model Comparisons

Model	Parameters	MSD ( $\times 100$ ) by data set		
		Devi	Rep	Rand
RELACS	$\lambda = 8, \alpha = 0.00125, \beta = 0.2, \kappa = 4$	0.36	0.40	0.53
Baseline				
Extended probability matching (EPM) (MSD for the Devi set computed on second block)	$b1 = 0.62, b2 = 0.22$	(0.50)	0.91	
Maximization and/or equilibrium <sup>a</sup>		7.92	9.02	1.12
Random		5.28	8.49	1.14
No-problem effect (the parameters are the blocks' means over problems)	$b1 = 0.65, b2 = 0.72, b3 = 0.73, b4 = .74, b5 = .88$	2.11		
One-strategy variants of RELACS				
Stochastic fictitious play (SFP): RELACS with the constraints that the slow best reply rule is used in all trials and $S(t) = 1$	$\lambda = 1, \alpha = 0.033$	1.89	1.87	0.82
SFP-pv: (with payoff variability effect): Like SFP without the constraint $S(t) = 1$	$\lambda = 3, \alpha = 0.01$	1.09	1.71	1.43
Reinforcement learning (RL): Like SFP with the additional constraint that propensities are updated only in trials in which the alternative was selected	$\lambda = 1, \alpha = 0.033$	2.13	1.43	0.82
RL-pv (with payoff variability effect): Like RL without the constraint $S(t) = 1$	$\lambda = 3, \alpha = 0.01$	1.03	0.67	1.43
Two-strategy variants of RELACS				
Without slow best reply	$\lambda = 3, \alpha = 0.0014, \beta = 0.02, \kappa = 4$	1.69	1.05	0.64
Without fast best reply	$\lambda = 8, \alpha = 0.002, \kappa = 4$	0.52	0.70	0.67
Without case based	$\lambda = 5, \alpha = 0.00125, \beta = 0.2$	0.68	0.57	0.54
Three-strategy variants of RELACS				
RELACS w/o learning among cognitive strategies (random choice among the three strategies)	$\lambda = 16, \alpha = 0.0006, \beta = 0.2, \kappa = 4$	0.36	0.42	0.53
RELACS with the constraint $S(t) = S(1)$ (with stable payoff variability effect)	$\lambda = 9, \alpha = 0.0011, \beta = 0.2, \kappa = 4$	0.38	0.41	0.56
RELACS with the constraint $\beta = 1$ (maximal recency)	$\lambda = 10, \alpha = 0.0033, \kappa = 4$	0.54	0.53	0.60
RELACS with the constraint $\kappa = t$	$\lambda = 7, \alpha = 0.0014, \beta = 0.2$	0.50	0.58	0.65
RELACS with the constraint $\kappa = \text{round}(1/\beta)$	$\lambda = 12, \alpha = 0.001, \beta = 0.2$	0.37	0.41	0.54
RELACS with the constraint $\alpha = \beta/t$	$\lambda = 16, \beta = 0.2, \kappa = 4$	0.40	0.60	0.54
RELACS with the constraint that the weighted values of the actions are updated only after selection.	$\lambda = 11, \alpha = 0.0011, \beta = 0.2, \kappa = 4$	0.40	0.41	0.51

Note. The parameters were estimated on the 40 problems used to demonstrate the deviations from maximization (the Devi set). The cross-validation focuses on the 27 problems representatively sampled by Myers et al. (1961; the Rep set), and the 10 randomly selected games studied by Erev et al. (2002; the Rand set). RELACS = reinforcement learning among cognitive strategies; MSD = mean squared deviation.  
<sup>a</sup> Maximization in the Devi and Rep sets and equilibrium in the Rand set.

assumption of a diminishing payoff variability effect: The constraint  $S(t) = S(1)$  increases the MSD score by less than 2%. These results are important because they show that some of the details assumed in RELACS are not necessary to capture the current data set. Yet we favor the complete version of RELACS, because it is easy to come up with situations in which the simplifications examined here lead to unreasonable predictions. A clarifying thought experiment involves the following problem:

*Problem 41 (complete feedback)*

H	1,000 with certainty
L	1,001 with $p = 0.9$ , 0 otherwise

The random choice version of RELACS implies at least 30% L choices (even after long experience); a similar prediction is made

under reinforcement learning among strategies with the constraint  $S(t) = S(1)$ . The complete version of RELACS, however, leads to the reasonable prediction that experience moves behavior toward maximization in Problem 41.

Additional results, summarized in the last section of Table 2, show weak sensitivity to the amount of information used to compute the implications of the different strategies.

### Generalization Test

Recall that the 40 experimental conditions studied above were selected to demonstrate robust deviations from maximization. Thus, the results might have limited implications. It is possible that the models that fit the data well are biased to address interesting

deviations that occur in a narrow set of situations but cannot predict behavior in less interesting but more common problems. Two generalization tests (see Busemeyer & Wang, 2000) were conducted to explore this possibility. The first focuses on a representative set of choice problems (see Gigerenzer et al., 1991) and the second on a random set of games.

*A Representative Sample of Probability-Learning Tasks*

The representative set of problems considered here was collected in Myers et al.'s (1961) systematic evaluation of the relative importance of cost, gain, and probabilities in a binary probability-learning paradigm. They examined the 27 pairs of gambles presented in Table 3 (with the conversion rate of half a cent for each chip). Eight DMs were presented with each pair of gambles for 150 trials. The DMs' goal was to predict which of two lights would be turned on, and the payoffs represented the outcomes of the possible contingencies. After each choice, one of the two lights was turned on (on the basis of the relevant probabilities).

The main conclusion of this elegant study was that choice probabilities are very sensitive to differences in expected values (DEV). In the last block of 50 trials (all the data presented by Myers et al., 1961), the correlation between the observed proportion of choices and the DEV was .94. This conclusion is reinforced with the observation that the stochastic fictitious play model provides a very good fit to this data set (MSD of 0.0043 with the parameters  $\alpha = .67$  and  $\lambda = 1$ , but MSD of 0.0187 with the

parameters estimated above). However, the large difference between the fit of SFP in the two data sets may not reflect a large difference in behavior. As shown in Table 2, RELACS and its variants considered here provide even better predictions of Myers et al.'s results on the basis of the parameters estimated in the first section. Over the 27 conditions, the correlation between the observed proportions and the RELACS predictions is .98 and the MSD score is 0.0040.

Katz (1964) replicated five of the conditions run by Myers et al. (1961). He used a similar paradigm, but the experiment lasted 400 trials. The column on the far right of Table 3 shows the estimated proportion of "left" choices between Trials 101 and 150 in Katz's study. (Katz presented the data in blocks of 100 trials. The estimate was computed as  $0.25[\text{first block}] + 0.75[\text{second block}]$ .) An examination of these values shows that in four out of the five cases, Katz's results are closer to RELACS than to Myers et al.'s results. Given that Myers et al. ran 8 subjects in each condition, this finding implies that in the current context RELACS is more accurate than a (50-trial) experiment with 8 subjects. These results support the optimistic hypothesis presented above. It seems that the effect of the three behavioral tendencies studied here is robust.

In addition to the good quantitative approximation, RELACS provides a simple verbal explanation for the good descriptive value of the EV rule in this 27-problem data set. Under RELACS, choice probability is expected to be sensitive to differences between the normalized payoffs,  $DEV/S(t)$ . In Myers et al. (1961), payoff

Table 3  
*The 27 Problems Studied by Myers et al. (1961)*

Problem	The gambles' parameters			Expected value (EV)			P(A) ( $\times 100$ )		
	P(E)	L	G	EV(A)	EV(B)	DEV	Observed	RELACS	Katz
1	0.5	-1	1	0.0	0.0	0.0	55	50	49
2	0.7	-1	1	0.4	-0.4	0.8	80	83	
3	0.9	-1	1	0.8	-0.8	1.6	96	96	
4	0.5	-1	2	0.0	0.5	-0.5	35	36	40
5	0.7	-1	2	0.4	-0.1	0.5	63	74	
6	0.9	-1	2	0.8	-0.7	1.5	96	95	
7	0.5	-1	4	0.0	1.5	-1.5	33	25	35
8	0.7	-1	4	0.4	0.5	-0.1	46	58	
9	0.9	-1	4	0.8	-0.5	1.3	86	92	
10	0.5	-2	1	-0.5	0.0	-0.5	28	36	39
11	0.7	-2	1	0.1	-0.4	0.5	76	74	
12	0.9	-2	1	0.7	-0.8	1.5	91	95	
13	0.5	-2	2	-0.5	0.5	-1.0	23	29	
14	0.7	-2	2	0.1	-0.1	0.2	60	66	
15	0.9	-2	2	0.7	-0.7	1.4	90	94	
16	0.5	-2	4	-0.5	1.5	-2.0	34	23	
17	0.7	-2	4	0.1	0.5	-0.4	54	52	
18	0.9	-2	4	0.7	-0.5	1.2	92	91	
19	0.5	-4	1	-1.5	0.0	-1.5	11	26	38
20	0.7	-4	1	-0.5	-0.4	-0.1	65	58	
21	0.9	-4	1	0.5	-0.8	1.3	91	92	
22	0.5	-4	2	-1.5	0.5	-2.0	18	23	
23	0.7	-4	2	-0.5	-0.1	-0.4	47	51	
24	0.9	-4	2	0.5	-0.7	1.2	89	91	
25	0.5	-4	4	-1.5	1.5	-3.0	16	21	
26	0.7	-4	4	-0.5	0.5	-1.0	33	42	
27	0.9	-4	4	0.5	-0.5	1.0	82	87	

Note. All the gamble pairs were of the following form: A (1 if E occurs, L otherwise) or B (G if E occurs, -1 otherwise). The right columns show the proportion of A choices in the last block of 50 trials (Trials 101-150) in Myers et al. (observed), according to RELACS's predictions, and in Katz (1964). DEV = differences in expected value; RELACS = reinforcement learning among cognitive strategies.

variance is generally stable across the 27 problems. Thus, correlation with  $DEV/S(t)$  implies correlation with  $DEV$  (e.g., the correlation between  $DEV$  and  $DEV/S(1)$  is .97). In the other data sets considered here and in other studies by Myers and his associates,  $S(t)$  varies from problem to problem. For example, in Myers et al. (1963, Problems 15–20), the correlation between  $DEV$  and  $DEV/S(1)$  is only .42. As a result, the correlation between  $DEV$  and choice probability over these problems is only .43.

### *Predicting Behavior in Games*

RELACS is an extension of models that were designed to capture learning in simple games (Erev et al., 1999; Erev & Roth, 1998; Roth & Erev, 1995). As Table 2 shows, RELACS outperforms the original models in summarizing the individual decision tasks studied here. To evaluate whether this advantage arises from differences between the types of tasks or better abstraction of the adaptation process, it is constructive to ask if RELACS, with the parameters estimated above, can capture the results of the simple games considered by the original models.

To answer this question, we derived the RELACS predictions for 10 randomly selected games with the mixed strategy equilibrium studied in Erev et al. (2002). Using the same unit of analysis as the previous research, we found that the MSD of RELACS in predicting the proportion of “left” choices in blocks of 100 trials is 0.0053. This score is similar to the MSD score of the basic reinforcement learning supported by the original article (and lower than the score of the basic model with the parameters estimated here). Erev et al. showed that the predictions of their model are as accurate as the predictions of an experiment with 9 pairs of subjects. RELACS is only slightly less useful. Its estimated *equivalent number of observations* (see Erev, Roth, Slonim, & Barron, 2002) in predicting choice behavior in blocks of 100 trials is larger than 7.

### *Model Comparison*

The columns on the right side of Table 2 compare the cross-validation MSD scores of the models considered here. The results reveal a similar advantage of RELACS and its close (three strategies) variants over the simpler models in the fitted and generalization sets. This pattern suggests that the observed advantage of the three-strategies abstraction documented above is not likely to result from data overfitting.

### *Limitations, Boundaries, and Related Phenomena*

Although the present research considers a relatively large set of data and with it we try to develop a relatively general model, the research has clear limitations and boundaries. The limitations arise from the focus on a particular statistic. Because the current model is only an approximated summary of the statistics we analyzed, it is unlikely that it will provide good predictions of statistics that are weakly correlated with the fitted statistics.

Boundaries come from the focus on the effect of immediate feedback on maximization rate in repeated binary decision tasks given a static environment. Many important learning phenomena occur outside this set of situations.

### *Limitations*

Human behavior is a product of the interactions between 100 billion neurons (Williams & Herrup, 1988) and many environmen-

tal variables. Thus, it is unlikely that a model with a small number of parameters will provide an accurate fit of observed behavior. Rather, models are summaries of particular sets of statistics, and it is often the case that different models best capture different statistics (Haruvy & Erev, 2001; see also Feltovich, 2000). To demonstrate the implications of this for the current analysis, we discuss four statistics that are not well captured by RELACS. Although RELACS can be extended to address these statistics, the current cost (in terms of the number of parameters relative to the number of observations) seems too high.

### *Individual Differences*

The white bars in Figure 5 show the observed distributions of  $P_{max}$  in the second block (Trials 101–200) over all the subjects in 32 of the conditions presented in Figures 2–4 (all the conditions for which individual data were available). The results reveal extremely large individual differences (and a similar pattern was observed in the other blocks). Indeed, a bimodal U-shaped distribution was observed in some of the problems. The light curves in Figure 5 display the distributions predicted by RELACS. Although RELACS predicts large individual differences, it clearly underpredicts the observed differences.

One version of RELACS that captures the observed individual differences involves a modified abstraction of the loss aversion and probability matching strategy. The original version assumes independence between the  $\kappa$  previous outcomes recalled by the DM in each trial. The modified abstraction relaxes this constraint and allows for the possibility that certain experiences are more memorable. For example, in one version, the first recalled outcome (used in the case-based strategy) is always the first outcome the DM observed during the experiment. This modification does not affect the fit of the aggregate curves (the MSD of this version of RELACS is 0.0036), but it increases the between-subjects variability to the level observed in the experiments.

### *Spontaneous Alternations and Sequential Dependencies*

While studying the behavior of rats in a simple T-maze task, Tolman (1925) discovered a robust violation of the law of effect. In each trial of his study, rats had to choose an arm in the T maze. Whereas the law of effect predicts an increase in choice probability after a reward, Tolman found a decrease. His subjects tended to alternate even after winning a reward in one of the two arms (for a review of this line of research, see Dember & Fowler, 1958).

Rapoport and Budescu (1992) have found a similar phenomenon in human behavior. In one of their experimental conditions, human subjects played a symmetrical zero-sum matching pennies game. Whereas their aggregate results (each alternative was chosen an almost equal number of times) are consistent with equilibrium and the RELACS predictions, analysis of sequential dependencies reveals a clear violation of these two models. Like Tolman’s (1925) rats, Rapoport and Budescu’s subjects exhibited a strong overalternation effect. In violation of the reinforcement learning prediction of a weak win–stay lose–change dependency, their subjects tended to alternate even after winning.

Another robust sequential dependency that violates the law of effect has been called the *gamblers’ fallacy* (or *negative recency*): At least in early stages of some probability learning studies (see Estes, 1964; Lee, 1971), DMs tend to predict a change in the state

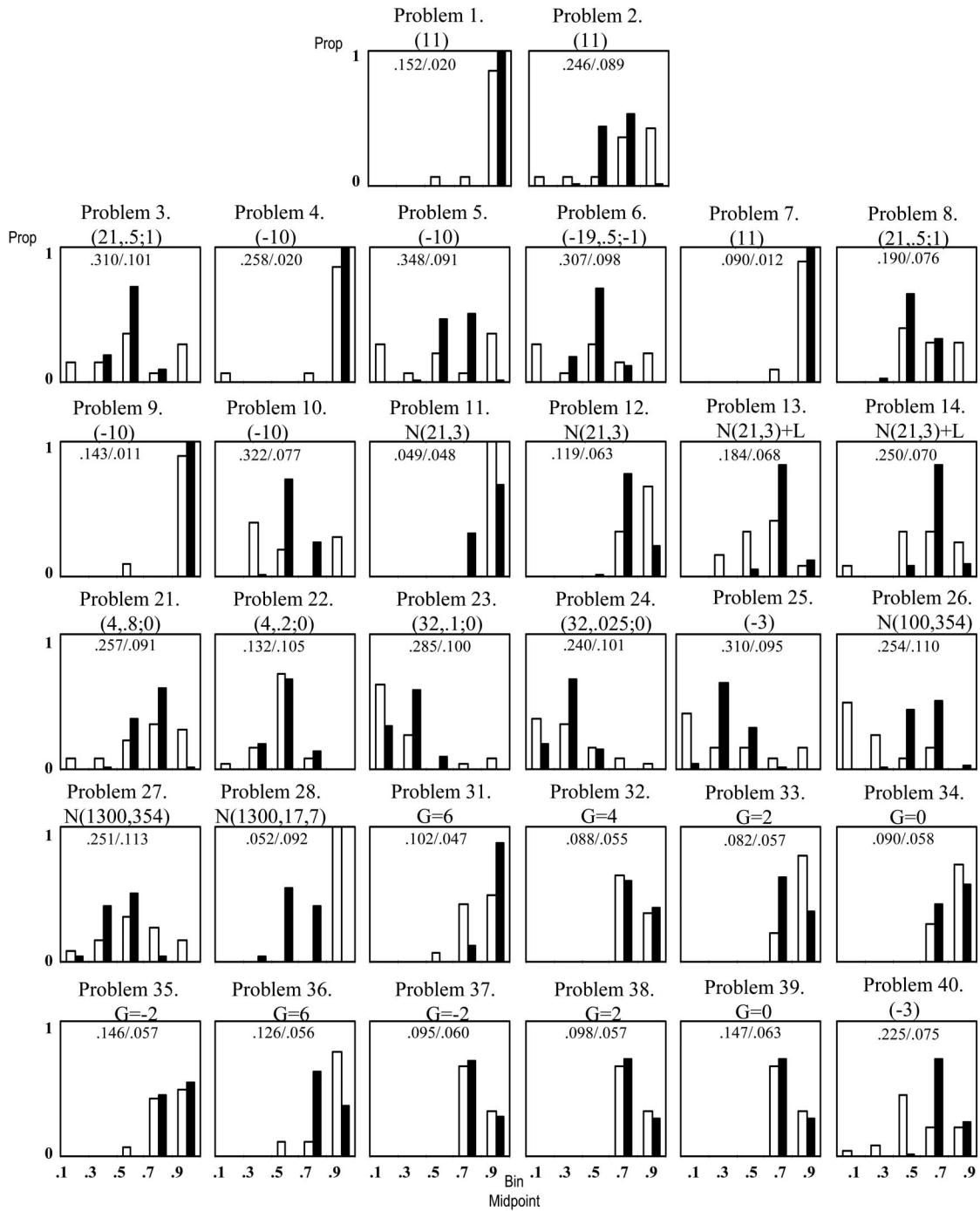


Figure 5. The distribution of the proportion of maximization in the second block (Pmax2) over decision makers (DMs) and simulated subjects in 32 binary choice problems. The x-axis shows the midpoint of Pmax2; the y-axis shows the proportion (Prop) of DMs (white bars) and simulated subjects (black bars) with this midpoint. The standard deviations for real and simulated subjects appear at the top of the graph. The graph titles denote the problem number and the gamble with higher expected values. N(x, y) means a draw from a normal distribution with mean x and standard deviation y.

of the world. Other studies (Bereby-Meyer, 1997) show large between-DM differences in sequential dependencies. Few DMs show an initial negative recency, and the large majority show initial positive recency.

The existence of negative recency and spontaneous alternations are clear violations of RELACS. Nevertheless, we do not believe they imply that the model is either not useful or should be immediately extended to account for these phenomena. Our belief is

based on the observation that the class of models considered here implies (in the current setting) a very weak relationship between the predicted maximization rate and the predicted sequential dependencies. (This assertion is not likely to hold in situations in which the environment is not static and maximization requires a response of sequential dependencies in the world. E.g., Green, Price, and Hamburger, 1995, showed that even pigeons can learn to maximize by alternation.) For example, the addition of an alternation or a negative recency strategy (see Rapoport, Erev, Abraham, & Olson, 1997) does not change the main results. This observation implies that the current goal (understanding maximization rate) can be achieved even without a good understanding of sequential dependencies. Rather, sequentially pursuing the two goals may be a good research strategy.

### *Predicting Trial $t + 1$ on the Basis of the First $t$ Trials*

Assume that you have observed the first 316 choices made by a particular DM in a particular choice problem and the obtained feedback. You are now asked to predict the DM's next choice (Trial 317). Should you use RELACS to make this prediction? The answer is clearly no. The large between-subjects variability presented in Figure 5 implies that a weighted average of the observed choices—for example, the single parameter model:  $P_j(t + 1) = w[P_j(t)] + (1 - w)[C_j(t)]$  where  $C_j(t) = 1$  if  $j$  was selected at  $t$  and 0 otherwise—will provide a better prediction.

The weighted average model is outperformed by models that were designed to predict trial  $t + 1$  on the basis of the first  $t$  (see, e.g., Camerer & Ho, 1999; Rapoport, Daniel, & Seale, 1998). Yet, like the weighted average model, these models do not provide a good summary of the current statistic. Indeed, they require situation-specific parameters and cannot be effectively used to predict behavior in new situations (see Haruvy & Erev, 2001; but see Rieskamp, Busemeyer, & Laine, 2003, and Yechiam & Busemeyer, 2005, for a demonstration of conditions under which the two methods provide converging results).

### *Predicting the Long Term*

Recall that the current analysis focused on experimental conditions that include a few hundreds trials. Can the model that best fits these data provide reliable predictions of behavior in the long term (e.g., after millions of trials)? We think that the answer is no. Our assertion is based on the observation (see Roth & Erev, 1995) that the long-term predictions of learning models can be highly sensitive to details that have little effect on the fit in the intermediate term (the observed experimental results). Thus, the current data cannot be used to decide between variants of RELACS that lead to very different predictions of the long term.

### *Boundaries*

To demonstrate the boundaries of the present research, we discuss in the current section four important learning regularities that are not addressed here.

### *Melioration and the Effect of Delayed Feedback*

Herrnstein and his associates (see a review in Herrnstein, Loewenstein, Prelec, & Vaughan, 1993) have demonstrated that in certain settings, experience leads DMs to meliorate (maximize

immediate payoffs) rather than to maximize long-term expected utilities. This observation is extremely important because there are many problems in which maximization of immediate rewards can impair long-term performance.

The focus of the current research on situations with immediate feedback implies that it cannot address these important findings. We chose not to address these findings here because the current understanding of the factors that mediate the effect of delayed outcomes is rather limited. For example, Herrnstein et al. (1993) found that a change in display can lead DMs toward maximization. Although they proposed a good qualitative summary of their findings, additional data have to be collected to allow the development of a general quantitative summary.

### *Multialternative Choice Tasks*

Study of repeated choice among a wide set of alternatives reveals a deviation from maximization that can be described as indicating a neighborhood effect: Decision makers tend to prefer alternatives that are displayed next to other relatively good alternatives (see Busemeyer & Myung, 1987; Rieskamp et al., 2003). It is easy to see that the current version of RELACS cannot capture this important pattern. One generalization of RELACS that captures the observed results assumes that when the strategy space is large (and there are no forgone payoffs), the fast best reply rule is generalized to imply a hill-climbing process (Yechiam, Erev, & Gopher, 2001).

### *Extinction, Transfer, and Humphreys's Paradox*

Humphreys (1939) noted that payoff variability slows the learning speed but leads to more robust learning. That is, it takes more time to extinguish behavior learned in a high-variability payoff condition. Grant, Hake, and Hornseth (1951) demonstrated this effect in the probability learning paradigm.

The common explanation of Humphreys's paradox assumes that extinction speed is sensitive to the organism's ability to detect a change in the environment (see, e.g., Mazur, 1994). This requires more time in a noisy environment. Thus, to capture this effect with the current model, the process of detecting change will have to be quantified. In addition, a precise assumption concerning the effect of the detected change on the propensities (e.g., fast forgetting and/or moving toward random choice) will be necessary (see Shor, 2002). The modeling of transfer from one environment to another might require additional modifications.

### *The Effect of Prior Information*

Recall that the current research focuses on situations in which DMs do not have prior information concerning the payoff distributions. This facilitates the examination of learning independent of the initial choice propensities and reduces the number of parameters. Obviously, this uniform initial assumption will have to be relaxed to address situations with known payoff distributions. One approach to this problem (see Barron, 2000) is to use a variant of prospect theory to capture the initial propensities. However, this solution is rather costly, as prospect theory has five free parameters.

### *Summary and Implications*

Early studies of decision making in iterated probability learning tasks revealed interesting deviations from maximization that can

be captured with the probability matching assumptions (see Estes, 1950; Grant et al., 1951). However, direct examination of this assumption shows that in many settings behavior is closer to maximization than to probability matching (see Edwards, 1961; Myers et al., 1961; Shanks et al., 2002; Siegel & Goldstein, 1959). Moreover, recent research that uses different experimental paradigms documents behavioral regularities that appear to be inconsistent with probability matching (see Barron & Erev, 2003). In the current review, we try to clarify this complex picture. We suggest that the availability of immediate feedback is not sufficient to lead choice behavior toward maximization (at least not during an experimental session that consists of 500 trials). The observed deviations from maximization can be attributed to three distinct behavioral effects. One set of deviations can be classified as indicating a payoff variability effect (Busemeyer & Townsend, 1993; Myers et al., 1961): An increase in payoff variability seems to move choice behavior toward random choice. A second set of deviations indicates underweighting of rare events (Barron & Erev, 2003): DMs tend to prefer the alternatives that lead to the best payoff most of the time even when those alternatives are associated with slightly lower expected return. A third set of deviations involves loss aversion (see Kahneman & Tversky, 1979): In certain cases, subjects tend to prefer alternatives that minimize losses over those that maximize payoffs.

The probability matching assumption (Estes, 1950) provides a useful approximation of the observed results. In most cases, the observed choice proportions deviate from maximization in the direction of matching the probability that the selected alternative will lead to the best outcomes. In the first 200 trials, the observed choice proportions are much closer to probability matching than to maximization. However, this approximation is biased in five ways. Four of the five biases can be addressed with an extended probability matching model that allows for matching samples (matching the probability that the selected alternative will produce the best outcome in samples of four observations), random choice, and loss aversion.

The main shortcomings of the extended probability matching model involve its failure to capture the dynamic process. This model ignores the observation that maximization tends to increase with experience, as well as the stickiness effect: the observation that the effect of negative experiences is longer lasting when feedback is limited to the obtained payoffs. These shortcomings are naturally addressed with a model, referred to as RELACS, that assumes there are reinforcement learning processes among three cognitive strategies. Although we selected the three cognitive strategies to fit the three deviations from maximization, there is no one-to-one correspondence between the assumed strategies and the observed deviations. Rather, under RELACS, the three deviations emerge as negative by-products of reasonable and well-known behavioral tendencies: fast best reply to recent outcomes (see Bush & Mosteller, 1955; Estes, 1950; March, 1996; Sarin & Vahid, 2001), loss aversion and case-based reasoning (see Gilboa & Schmeidler, 1995; Kahneman & Tversky, 1979; Logan, 1988; Riesbeck & Schank, 1989), and slow best reply with exploration (see Busemeyer, 1985; Erev et al., 1999; Friedman & Mezzetti, 2001; Myers & Sadler, 1960; Weber et al., 2004).

The value of RELACS was demonstrated in five analyses. The first showed that with a single set of four parameters, RELACS provides a good quantitative fit of the observed

maximization rates in all 40 problems reviewed here. A second analysis revealed that it is not easy to find a simpler model with similar descriptive value and that all the assumptions made in RELACS appear to be necessary. A third (sensitivity) analysis suggested that the good fit provided by RELACS is not a result of overfitting the data with the assumptions implicit in the quantification of the different principles. The good fit is not sensitive to the details of the quantification. A fourth analysis illustrated the high predictive power of RELACS in forecasting behavior in the 27 tasks studied by Myers et al. (1961; a correlation of .98). A fifth analysis showed generality to a set of randomly selected matrix games.

This “three reasonable strategies” summary of the experimental findings supports two suggestions. The first concerns the robustness of the observed deviations from maximization. The fact that the deviations occur naturally as likely by-products of reasonable and well-known tendencies suggests that they do not reflect rare exceptions that can be ignored. It seems that in the settings examined here (adaptation given limited information), ecologically reasonable behavioral tendencies do not ensure maximization. Moreover, the deviations from maximization appear to be stable even after hundreds of trials with immediate feedback.

The second suggestion has to do with the relation between the behavioral tendencies documented here (in iterated choice settings) and those documented in studies of decision making in one-shot tasks (e.g., the research summarized by prospect theory; Kahneman & Tversky, 1979). Although DMs tend to deviate from maximization in both settings, in three cases, the direction of the deviations can be reversed. In particular, (a) DMs behave as if they overweight rare outcomes in one-shot tasks but underweight them in iterated tasks; (b) outcomes that are provided with certainty are overweighted in one-shot tasks but can be underweighted in iterated tasks; and (c) in certain problems, DMs show the reflection effect (risk aversion in the gain domain and risk seeking in the loss) in one-shot tasks but the reverse pattern in iterated tasks. These large differences imply that March’s (1996) assertion that adaptation can lead to the behavior predicted by prospect theory is too optimistic. Different psychological processes seem to inform behavior in the two cases. Thus, additional experimental research is needed to highlight the factors that trigger the different processes.

#### *Relationship to Decision Field Theory (DFT)*

The choice rule described in Equation 2 is closely related to the choice rule assumed by DFT (Busemeyer & Townsend, 1993; Roe, Busemeyer, & Townsend, 2001). DFT’s choice rule (the random subjective expected utility rule initially proposed by Busemeyer, 1985) can be summarized by Equation 2 with  $q_j(t)$  as the expected utility from selecting  $j$  and  $S(t)$  as the standard deviation of the payoff differences. This similarity implies that in many situations, RELACS and DFT yield similar predictions. One example involves the data presented in Table 2. The correlation between the predictions of RELACS and DFT (Busemeyer & Townsend, 1993, estimated the parameters of DFT on the basis of 29 similar conditions run by Myers, Suydam, & Gambino, 1965, and Katz, 1964) for this data set is .995. The MSD of DFT is 0.0039, slightly better than RELACS (MSD = 0.0040).

This similarity, however, is not general. There are important differences between RELACS and DFT. DFT was proposed to

capture choice probability and reaction time. It does not predict the learning curve (DFT's predictions can be thought of as the result of experience), and it assumes choice among actions (rather than cognitive strategies). This observation suggests that when cognitive strategies are important, RELACS and DFT make different predictions. Indeed, DFT does not predict the underweighting of rare events. And additional parameters have to be added to DFT to capture loss aversion.

Yet the apparent advantage of RELACS over DFT in predicting choice probabilities does not imply that RELACS is an alternative to DFT. DFT offers useful insights into many phenomena that are outside the scope of RELACS. We therefore hope to examine, in future research, the possibility of combining DFT and RELACS in a single model that predicts RELACS-like adaptation and DFT-like deliberation and reaction time.

### Practical Implications

The current review implies that relatively small modifications of the available feedback can have large and predictable effects on expected behavior. This observation has many non-trivial implications. One set of implications involves manipulations that eliminate the risk that DMs will underweight rare events. For example, consider the design of red light cameras. Typically, these cameras operate (and the cameras' flash can be observed) when a car enters the junction more than 1 s into the red light. Thus, the typical feedback from running a red light is positive: The driver can see that he or she was not detected. As a result, underweighting of rare events is expected to increase the tendency to run red lights. To address this problem, Perry, Haruvy, and Erev (2002) proposed to reduce the delay between the beginning of the red light and the operation of the cameras' flash. On the basis of the current analysis, they show that this change can be effective even if the probability of a fine given a flash is not high (and the change does not modify the relevant expected payoffs).

Other applications include methods to reduce payoff variance, such as the auction mechanism proposed by Erev et al. (2004). They show that this mechanism facilitates efficiency when incorporated in new flight regulations designed to determine right-of-way when two aircraft meet in midair.

Obviously, the current results can also be used to take advantage of naive agents. Many of these problematic implications are already in use. Indeed, the design of casino slot machines (e.g., high payoff variability that impairs learning) is one example. We hope that the current explicit presentation of these principles will help reduce this risk.

### References

- Allais, M. (1979). The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school. In M. Allais & O. Hagen (Eds.), *Expected utility hypotheses and the Allais paradox: Contemporary discussions of decision under uncertainty with Allais' rejoinder* (pp. 27–145). Dordrecht, the Netherlands: D. Reidel.
- Barron, G. (2000). *The effect of feedback on decision making under uncertainty and risk: The predictive value of descriptive models*. Unpublished master's thesis, Technion-Israel Institute of Technology, Haifa, Israel.
- Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making, 16*, 215–233.
- Bereby-Meyer, Y. (1997). *On the robustness of the framing effect: A re-interpretation of the "payoff effect" in probability learning experiments*. Unpublished doctoral dissertation, Technion-Israel Institute of Technology, Haifa, Israel.
- Bereby-Meyer, Y., & Erev, I. (1998). On learning to become a successful loser: A comparison of alternative abstraction of learning processes in the loss domain. *Journal of Mathematical Psychology, 42*, 266–286.
- Berry, D., & Fristedt, B. (1985). *Bandit problems: Sequential allocation of experiments*. London: Chapman & Hall.
- Blavatskyy, P. R. (2004). *Essays on cumulative prospect theory, individuals' preference for most probable winner, and the design of Olympic prizes*. Unpublished doctoral dissertation, Charles University, Prague, Czech Republic.
- Busemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed sample, and sequential sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 538–564.
- Busemeyer, J. R., & Myung, I. J. (1987). Resource allocation decision making in an uncertain environment. *Acta Psychologica, 66*, 1–19.
- 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., & 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.
- Bush, R., & Mosteller, F. (1955). *Stochastic models for learning*. New York: Wiley.
- Camerer, C. F., & Ho, T. (1999). Experience-weighted attraction in games. *Econometrica, 64*, 827–874.
- Cheung, Y. W., & Friedman, D. (1998). A comparison of learning and replicator dynamics using experimental data. *Journal of Economic Behavior and Organizations, 35*, 263–280.
- Dember, W. N., & Fowler, F. (1958). Spontaneous alternation behavior. *Psychological Bulletin, 55*, 412–428.
- Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science, 12*, 523–538.
- Diederich, A., & Busemeyer, J. R. (1999). Conflict and the stochastic dominance principle of decision making. *Psychological Science, 10*, 353–359.
- Edwards, W. (1961). Probability learning in 1000 trials. *Journal of Experimental Psychology, 62*, 385–394.
- Erev, I. (1994). Convergence in the orange grove: Learning processes in a social dilemma setting. In U. Schul, W. Albers, & U. Mueller (Eds.), *Social dilemmas and cooperation* (pp. 187–206). New York: Springer-Verlag.
- 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., Barron, G., & Remington, R. (2004). Right of way in the sky: Two problems in aircraft self-separation and the auction-based solution. *Human Factors, 46*, 277–287.
- Erev, I., Bereby-Meyer, Y., & Roth, A. E. (1999). The effect of adding a constant to all payoffs: Experimental investigation, and implications for reinforcement learning models. *Journal of Economic Behavior & Organization, 39*, 111–128.
- Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in games with unique strategy equilibrium. *American Economic Review, 88*, 848–881.
- Erev, I., & Roth, A. E. (1999). On the role of reinforcement learning in experimental games: The cognitive game-theoretic approach. In D.

- Budescu, I. Erev, & R. Zwick (Eds.), *Games and human behavior: Essays in honor of Amnon Rapoport* (pp. 53–78). Mahwah, NJ: Erlbaum.
- Erev, I., Roth, A. E., Slonim, S. L., & Barron, G. (2002). Predictive value and the usefulness of game theoretic models. *International Journal of Forecasting*, *18*, 359–368.
- Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*, 94–107.
- Estes, W. K. (1964). Probability learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 89–128). New York: Academic Press.
- Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, *83*, 37–64.
- Feltoch, N. (2000). Reinforcement-based vs. belief-based learning models in experimental asymmetric-information games. *Econometrica*, *68*, 605–641.
- Fiedler, K. (2000). Beware of samples! A cognitive–ecological sampling approach to judgment biases. *Psychological Review*, *107*, 659–676.
- Friedman, J. W., & Mezzetti, C. (2001). Learning in games by random sampling. *Journal of Economic Theory*, *98*, 55–84.
- Fudenberg, D., & Levine, D. (1998). *Theory of learning in games*. Cambridge, MA: MIT Press.
- Gigerenzer, G., Hoffrage, U., & Kleinulting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.
- Gilboa, I., & Schmeidler, D. (1995). Case-based decision theory. *The Quarterly Journal of Economics*, *110*, 605–639.
- Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, *112*, 631–645.
- Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, *38*, 129–166.
- Grant, D., Hake, H. W., & Hornsby, J. P. (1951). Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *Journal of Experimental Psychology*, *42*, 1–5.
- Green, L., Price, P. C., & Hamburger, M. E. (1995). Prisoner's dilemma and the pigeon: Control by immediate consequences. *Journal of the Experimental Analysis of Behavior*, *64*, 1–17.
- Grosskopf, B., Erev, I., & Yechiam, E. (2005). *Forgone with the wind*. Manuscript submitted for publication.
- Haruvy, E., & Erev, I. (2001). On the application and interpretation of learning models. In R. Zwick & A. Rapoport (Eds.), *Advances in experimental business research* (pp. 285–300). Norwell, MA: Kluwer Academic.
- Haruvy, E., & Erev, I. (2002). *Variable pricing: A customer learning perspective*. Unpublished manuscript, Technion–Israel Institute of Technology, Haifa, Israel.
- Haruvy, E., Erev, I., & Sonsino, D. (2001). The medium prizes paradox: Evidence from a simulated casino. *Journal of Risk and Uncertainty*, *22*, 251–261.
- Herrnstein, R. J., Loewenstein, G. F., Prelec, D., & Vaughan, W., Jr. (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioral Decision Making*, *6*, 149–185.
- Hertwig, R., & Ortman, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, *24*, 383–451.
- Humphreys, L. G. (1939). The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *Journal of Experimental Psychology*, *25*, 141–158.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.
- Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review*, *107*, 397–402.
- Katz, L. (1964). Effects of differential monetary gain and loss on sequential two-choice behavior. *Journal of Experimental Psychology*, *68*, 245–249.
- Lee, W. (1971). *Decision theory and human behavior*. New York: Wiley.
- Logan, G. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3, pp. 249–410). New York: Wiley.
- March, J. G. (1996). Learning to be risk averse. *Psychological Review*, *103*, 309–319.
- Mazur, J. E. (1994). *Learning and behavior*. Englewood Cliffs, NJ: Prentice Hall.
- Myers, J. L., 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.
- Myers, J. L., Reilly, E., & Taub, H. A. (1961). Differential cost, gain, and relative frequency of reward in a sequential choice situation. *Journal of Experimental Psychology*, *62*, 357–360.
- Myers, J. L., & Sadler, E. (1960). Effects of range of payoffs as a variable in risk taking. *Journal of Experimental Psychology*, *60*, 306–309.
- Myers, J. L., Suydam, M. M., & Gambino, B. (1965). Contingent gains and losses in risk-taking situations. *Journal of Mathematical Psychology*, *2*, 363–370.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, United Kingdom: Cambridge University Press.
- Perry, O., Haruvy, E., & Erev, I. (2002). Frequent delayed probabilistic punishment in law enforcement. *Economics of Governance*, *3*, 71–85.
- Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General*, *121*, 352–363.
- Rapoport, A., Daniel, T. E., & Seale, D. A. (1998). Reinforcement-based adaptive learning in asymmetric two-person bargaining with incomplete information. *Experimental Economics*, *1*, 221–253.
- 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.
- Riesbeck, C. K., & Schank, R. (1989). *Inside case-based reasoning*. Northvale, NJ: Erlbaum.
- Rieskamp, J., Bussemeyer, J., & Laine, T. (2003). How do people learn to allocate resources? Comparing two learning theories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1066–1081.
- Roe, R. M., Bussemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, *108*, 370–392.
- Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, *8*, 164–212.
- Sarin, R., & Vahid, F. (2001). Predicting how people play games: A simple dynamic model of choice. *Games and Economic Behavior*, *34*, 104–122.
- Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*, 233–250.
- Shor, M. (2002). *Learning to respond: The use of heuristics in dynamic games*. Unpublished manuscript, Vanderbilt University, Nashville, TN.
- Siegel, S., & Goldstein, D. A. (1959). Decision-making behavior in a two-choice uncertain outcome situation. *Journal of Experimental Psychology*, *57*, 37–42.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Stahl, D. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior*, *16*, 303–330.
- Thaler, R., Tversky, A., Kahneman, D., & Schwartz, A. (1997). The effect of myopia and loss aversion on risk taking: An experimental test. *Quarterly Journal of Economics*, *112*, 647–661.
- Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Monographs*, *2*(4).
- Tolman, E. C. (1925). Purpose and cognition: The determiners of animal learning. *Psychological Review*, *32*, 285–297.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory:

- Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 9, 195–230.
- von Neumann, J., & Morgenstern, O. (1947). *The theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
- Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, 14, 101–118.
- Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, 111, 430–445.
- Williams, R. W., & Herrup, K. (1988). The control of neuron number. *Annual Review of Neuroscience*, 11, 423–453.
- Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience based decision-making. *Psychonomic Bulletin & Review*, 12, 387–402.
- Yechiam, E., Erev, I., & Gopher, D. (2001). On the potential value and limitations of emphasis change and other exploration-enhancing training methods. *Journal of Experimental Psychology: Applied*, 7, 277–285.

Received February 19, 2004

Revision received February 28, 2005

Accepted March 22, 2005 ■

### **New Editor Appointed, 2007–2012**

The Publications and Communications (P&C) Board of the American Psychological Association announces the appointment of a new editor for a 6-year term beginning in 2007. As of January 1, 2006, manuscripts should be directed as follows:

- *Emotion* ([www.apa.org/journals/emo.html](http://www.apa.org/journals/emo.html)), **Elizabeth A. Phelps, PhD**, Department of Psychology, New York University, 6 Washington Place, Room 863, New York, NY 10003.

**Electronic manuscript submission.** As of January 1, 2006, manuscripts should be submitted electronically via the journal's Manuscript Submission Portal (see the Web site listed above). Authors who are unable to do so should correspond with the editor's office about alternatives.

Manuscript submission patterns make the precise date of completion of the 2006 volumes uncertain. The current editors, Richard J. Davidson, PhD, and Klaus R. Scherer, PhD, will receive and consider manuscripts through December 31, 2005. Should 2006 volumes be completed before that date, manuscripts will be redirected to the new editor for consideration in 2007 volume.