Cognitive Science 32 (2008) 1376–1402

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1080/03640210802352992

# Comparison of Decision Learning Models Using the Generalization Criterion Method

Woo-Young Ahn[a], Jerome R. Busemeyer[a], Eric-Jan Wagenmakers[b], Julie C. Stout[a,c]

[a]*Department of Psychological and Brain Sciences, Indiana University*
[b]*Department of Psychology, University of Amsterdam*
[c]*School of Psychology, Psychiatry, and Psychological Medicine, Monash University, Australia*

Received 11 October 2007; received in revised form 9 April 2008; accepted 18 July 2008

**Abstract**

It is a hallmark of a good model to make accurate *a priori* predictions to new conditions (Busemeyer & Wang, 2000). This study compared 8 decision learning models with respect to their generalizability. Participants performed 2 tasks (the Iowa Gambling Task and the Soochow Gambling Task), and each model made a priori predictions by estimating the parameters for each participant from 1 task and using those same parameters to predict on the other task. Three methods were used to evaluate the models at the individual level of analysis. The first method used a post hoc fit criterion, the second method used a generalization criterion for short-term predictions, and the third method again used a generalization criterion for long-term predictions. The results suggest that the models with the prospect utility function can make generalizable predictions to new conditions, and different learning models are needed for making short- versus long-term predictions on simple gambling tasks.

*Keywords:* Generalization criterion; Reinforcement learning; Decision making

## 1. Introduction

The study of human cognition has greatly profited from the development and application of quantitative models, including both mathematical and computer simulation models. These models have the potential to provide a coherent theoretical explanation of otherwise isolated empirical effects, to offer precise predictions instead of vague verbalizations, and to yield novel insights that challenge and replace old ideas once thought unassailable. An additional advantage is the opportunity to use quantitative methods to compare and evaluate competing models, which is critical for making scientific progress in cognition.

Correspondence should be sent to Woo-Young Ahn, Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405. E-mail: ahnw@indiana.edu

One of the main difficulties with quantitative comparisons among models is the problem of model complexity. A complex model may provide a more accurate post hoc fit to a particular data set than a simpler model, although the simpler model may provide more accurate *a priori* predictions for new conditions, tasks, or experiments. Shiffrin, Lee, Wagenmakers, and Kim (in press) provided a general overview of various methods to deal with this problem when performing model comparisons. The purpose of this article is to present a detailed application of one particular method, the *generalization criterion* (Busemeyer & Wang, 2000). Here we use this method to compare different models of decision and learning (e.g., Yechiam & Busemeyer, 2005) on two experience-based, decision-making tasks, the Iowa Gambling Task (IGT; Bechara, Damasio, Tranel, & Anderson, 1994) and the Soochow Gambling Task (SGT; Chiu et al., 2008).

The generalization criterion for model comparison is based on the following simple to use procedure: (a) A set of experimental conditions is divided into two subsets—a calibration set and a test set, (b) all of the data obtained from the calibration set of conditions are used to estimate the free parameters of the models, (c) the parameters estimated from the calibration stage are then used to generate new a priori predictions for the data obtained from the set of test conditions, and (d) the predictive accuracies of the models under the test conditions are used to compare and evaluate the competing models.

The generalization method differs from the cross-validation method in the following important way: Cross-validation randomly divides the entire set of data from all conditions into two subsets of data—a calibration set of data points and a test set of data points. Thus, data are sampled from all the conditions for the calibration stage. Generalization explicitly restricts sampling of data points for the calibration stage to a subset of conditions, and so parameters are not fit to any data from the test conditions. Cross-validation is useful for small sample sizes, but it favors more complex models (whether true or not) with large sample sizes (Browne, 2000). In contrast, the generalization criterion can be used to select a simpler model that generalizes better than a more complex model, even with very large sample sizes (Busemeyer & Wang, 2000; Forster, 2000).

The generalization criterion has been used in several previous applications (e.g., Erev & Barron, 2005; Erev & Haruvy, 2005; Rieskamp, Busemeyer, & Laine, 2003; Yechiam & Busemeyer, 2005), but these studies were based on comparisons across different groups of participants. The average of parameters obtained from a calibration group were used to generate predictions for a generalization test group. This study used a more powerful within-subjects design that enabled the application of the generalization criterion at an individual level of analysis (Yechiam & Busemeyer, 2008). Each participant performed two multi-alternative gambling tasks with order counterbalanced across participants. This within-subjects design allowed us to post hoc fit the model parameters to one of the gambling tasks for an individual (e.g., fit choices on the first task), and then use these parameter estimates to make a priori predictions for this person on the other gambling task (e.g., predict choices on the second task).

The remainder of this article is organized as follows. First, we describe the experimental procedures and the basic findings. Next, we describe the model fits and the model comparisons based on the generalization criterion. Finally, we describe implications of these findings for

the comparison of learning theories and the utility of the generalization method for comparing models more generally.

## 2. Experiments and basic findings

In this study, 2 four-alternative simulated gambling tasks were used: the IGT and the SGT. In both of the gambling tasks, participants were presented with four decks (*or* alternatives) and were asked to choose a card from any of the four decks. Every selection yields a gain or loss of some amount, and the goal of both tasks is to make as much money as possible. Fig. 1 shows a screenshot of the IGT as an example.

The IGT and the SGT differ in how payoffs are presented and in their payoff distributions (see Tables 1 and 2). In the IGT, gains and losses are presented to participants separately (e.g., a participant learns that she or he won \$1.00 but lost \$2.00), whereas in the SGT only the net gains are presented to participants (e.g., a participant is told that she or he lost \$1.00). Tables 1 and 2 illustrate the payoff distributions of the IGT and the SGT, respectively. In both tasks, Decks A and B are disadvantageous (or *bad*) with regard to their long-term gain (expected value of 10 trials = −\$2.5), and Decks C and D are advantageous (or *good*) with positive long-term gain (expected value of 10 trials = \$2.5). The location of these four decks are randomized for each participant, and decks are unlabeled when presented to participants. Choosing the advantageous decks maximizes expected value.

From a statistical perspective, the IGT and the SGT are so-called four-armed bandit problems (Berry & Fristedt, 1985). Bandit problems are a special case of the more general reinforcement learning problems in which an agent has to learn an environment by choosing actions and experiencing the consequences of those actions (e.g., Estes, 1950; Sutton & Barto, 1998). Optimal performance on such problems depends on a delicate trade-off between


Fig. 1. A sample display of the Iowa Gambling Task.

Table 1
The payoff distribution of the Iowa Gambling Task for the first 10 trials

| Deck | A | B | C | D |
|---|---|---|---|---|
| Expected value of 10 trials ($) | −2.50 | −2.50 | 2.50 | 2.50 |
| Gain on every trial ($) | +1.00 | +1.00 | +0.50 | +0.50 |
| Loss on each trial | | | | |
| Trial 1 | | | | |
| Trial 2 | | | | |
| Trial 3 | −1.50 | | −0.50 | |
| Trial 4 | | | | |
| Trial 5 | −3.00 | | −0.75 | |
| Trial 6 | | | | |
| Trial 7 | −2.00 | | −0.25 | |
| Trial 8 | | | | |
| Trial 9 | −2.50 | −12.50 | −0.50 | |
| Trial 10 | −3.50 | | −0.50 | −2.50 |

*Note*. These 10 trials were repeated 12 times ($10 \times 12 = 120$ trials). The order of these 10 trials was randomized every time.

"exploration" and "exploitation"; to discover the best option, the agent first has to try out or explore the various opportunities. However, if the agent only has a limited number of trials left, it is optimal to gradually stop exploring and instead exploit the option that has turned out to produce the highest profit in the past.

Both tasks have interesting characteristics in the context of human judgment and decision making. The IGT has been widely used to measure deficits in decision making exhibited by clinical populations, such as ventromedial prefrontal cortex (vmPFC) patients (Bechara et al., 1994, Bechara, Damasio, Tranel, & Damasio, 1997; Damasio, 1994) and drug abusers (for a review, see Yechiam, Stout, Busemeyer, Rock, & Finn, 2005). Numerous studies have shown

Table 2
The payoff distribution of the Soochow Gambling Task for the first five trials

| Deck | A | B | C | D |
|---|---|---|---|---|
| Expected value of 10 trials ($) | −2.50 | −2.50 | 2.50 | 2.50 |
| Net gain on each trial | | | | |
| Trial 1 | 1.00 | 0.50 | −1.00 | −0.50 |
| Trial 2 | 1.00 | 0.50 | −1.00 | −0.50 |
| Trial 3 | 1.00 | 0.50 | −1.00 | −0.50 |
| Trial 4 | 1.00 | 0.50 | −1.00 | −0.50 |
| Trial 5 | −5.25 | −3.25 | 5.25 | 3.25 |

*Note*. These five trials were repeated 24 times ($5 \times 24 = 120$ trials). The order of these five trials was randomized every time. Note that all decks of the Soochow Gambling Task have the same expected values as those of the Iowa Gambling Task.

that, whereas normal decision makers learn to choose more cards from advantageous decks, many patient groups fail to do so.

Bechara et al. (1997) proposed the Somatic Marker Hypothesis and claimed that normal decision makers are influenced by the longer term implications of their decisions, whereas patients with vmPFC damage make decisions using only the more immediate results of their actions. In decision-making terms, this claim is that normal decision makers choose more from Decks C and D because they have positive expected values, and fewer from Decks A and B because they have negative expected values. However, as Chiu et al. (2008) noted, selections from Decks C and D also have a higher frequency of net gains in comparison to Decks A and B (see Table 1); thus, expected values are confounded with gain frequency, and it is difficult to know which of these factors is responsible for the choice patterns observed in the IGT. Chiu et al. (in press) argued that perhaps choice behavior in the IGT is guided by the probability of net gain payoffs, rather than the expected values. To test this possibility, they developed the SGT, which instead links the advantageous decks (those with positive expected values) with low net gain frequencies (1 out of 5) and disadvantageous decks with high net gain frequencies (4 out of 5) while maintaining identical expected values to those in the IGT. Consistent with the interpretation that gain frequency rather than expected payoffs drives choice behavior, Chiu et al. (in press) found that healthy control participants preferred disadvantageous decks in the SGT. From the point view of Erev and Barron (2005), participants' choices deviate from advantageous decks in the SGT because of the normal tendency of decision makers to *underweight rare events*, to be *loss averse*, or both. Underweighting of rare events refers to the tendency to underweight small probabilities of rare events and prefer the alternative that provides the best outcome most of the time. Loss aversion refers to the tendency of participants to prefer the alternative that decreases the probability of losses even if it is associated with a lower expected gain.

Studies using the IGT indicate that normal healthy adults, on average, gradually learn which decks are advantageous for maximizing monetary gain; this learning is indicated by the observation that as trials proceed, they make an increasing proportion of choices from the advantageous decks. In contrast, in the SGT, although advantageous decks have the same average expected values as those in the IGT, normal decision makers (young healthy adults) were unable to learn which decks are good decks and, therefore, lost money (Chiu et al., 2008). In this study, the same data patterns were replicated using a within-subjects design, and the data sets were used to further evaluate decision learning models. The very different choice patterns in the two tasks allowed us to perform a critical generalization test because good fits to the pattern for the calibration data set do not guarantee good predictions to the pattern for the test data set (Busemeyer & Wang, 2000).

### *2.1. Methods*

#### *2.1.1. Participants*

Thirty-six individuals (18 men, mean age = 22.0, range = 18–33) from Indiana University, Bloomington campus area were recruited in the study. Most of them (31 out of 36) were

college students. They were paid $7 per hour and received bonuses, the amount of which depended on their performance on the tasks; the average bonus was $20.[1]

### 2.1.2. Procedures

Participants listened to instructions read aloud. Participants assigned to Group A first participated in the IGT and then the SGT. Participants in Group B completed the tasks in reverse order. The participants were randomly assigned to either group, and there were an equal number of men and women in each group. Each task had 120 trials, and the number of cards in each deck was unlimited for both tasks.

## 2.2. Results

The proportions of choices from advantageous decks on both tasks are plotted in Fig. 2. For data analysis, we divided 120 trials into six blocks of 20 trials and examined proportions of advantageous choices on each block. In the IGT, participants chose more cards from advantageous decks (Decks C and D) as trials went on, a result that is consistent with past research. An analysis of variance (ANOVA) revealed a main effect of block, $F(5, 175) = 14.0$, $p < .0001$, Greenhouse–Geisser corrected. Except for the first two blocks, proportions of choices from advantageous decks were significantly greater than 0.5 (one-sample $t$ test, $p < .001$ for Blocks 3–6). In contrast, for the SGT, participants chose fewer cards from

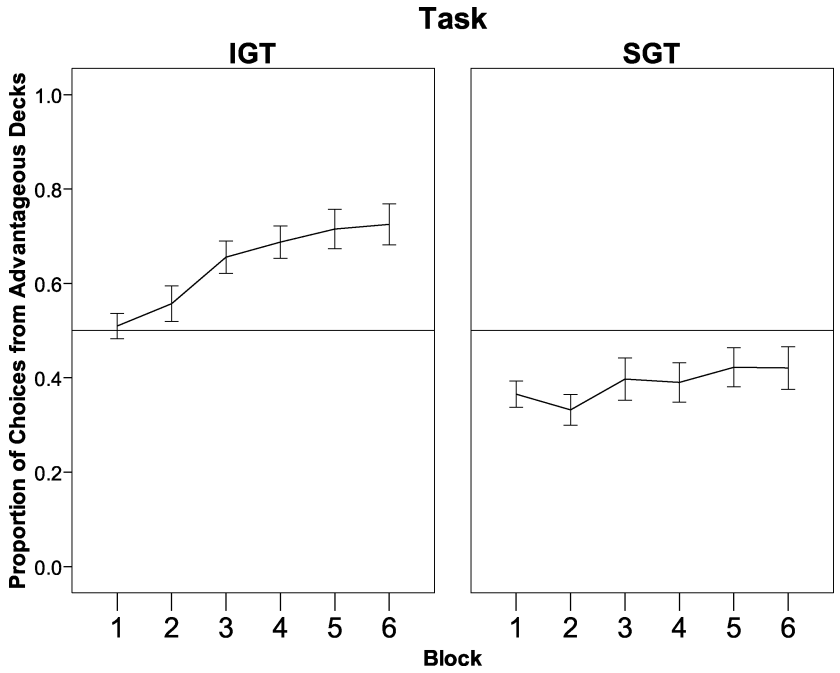


Fig. 2. Proportion of choices from advantageous decks in the Iowa Gambling Task (IGT) and the Soochow Gambling Task (SGT). Each block consists of 20 trials. Error bars indicate ±1 SEM.

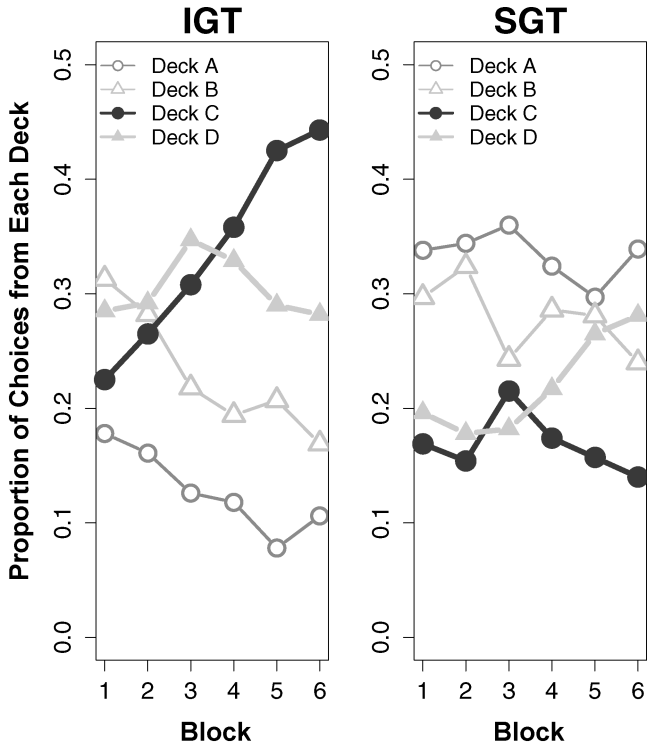


Fig. 3. The proportion of choices from each deck on the Iowa Gambling Task (IGT) and the Soochow Gambling Task (SGT). Each block consists of 20 trials.

advantageous decks, although the decks had the same expected values as those in the IGT. Except for Blocks 5 and 6, participants consistently chose more cards from disadvantageous decks (one-sample $t$ test, $p < .001$ for Blocks 1 and 2 and $p < .03$ for Blocks 3 and 4). A main effect of block was not significant for the SGT, $F(5, 175) = 1.5$, $p > .2$, Greenhouse–Geisser corrected. One possibility is that 120 trials could be too few, and participants in the SGT would ultimately learn to prefer the advantageous decks if the number of trials increases. Chiu, Lin, Huang, Lin, and Huang (2006) showed that even if the number of trials was increased to 200 trials, participants, on average, still consistently chose more cards from disadvantageous decks with little learning in later blocks. When we used parameters estimated from 120 trials of either the IGT or the SGT and ran simulations for 200 trials on the SGT, we had almost exactly the same pattern.[2]

The proportion of choices (pooled across all trials) from each of the four decks are shown in Fig. 3. Previous studies showed that, on average, decks with low-frequency negative payoffs (e.g., Deck B in the IGT) are chosen more often than high-frequency decks (Deck A in the IGT; Barren & Erev, 2003; Erev & Barron, 2005; Hertwig, Barren, Weber, & Erev, 2004; Yechiam & Busemeyer, 2005), even if the decks' expected payoffs are equal. Consistent with these past results, in the IGT of this study, Deck B, a disadvantageous alternative with low-frequency negative payoffs, was more frequently chosen than Deck A, a disadvantageous alternative

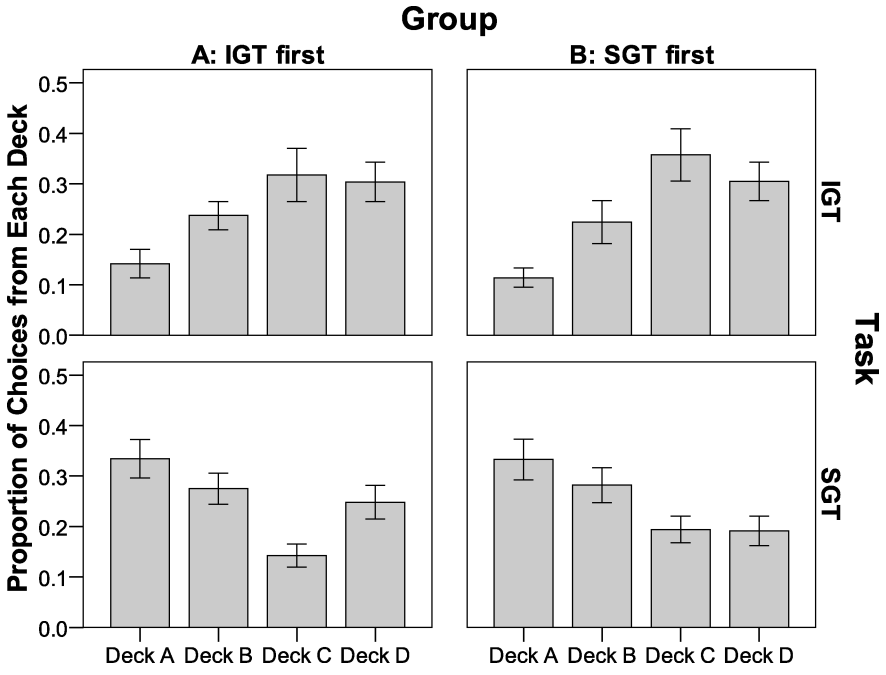


Fig. 4. Comparison of mean proportions of choices from each deck by group and task. Error bars indicate ±1 SEM. IGT = Iowa Gambling Task; SGT = Soochow Gambling Task.

with high-frequency negative payoffs, $t(35) = -3.35$, $p < .002$. However, the proportions of choices from Decks C and D were not significantly different, $t(35) = 0.59$, $ns$.[3]

The effects of task order are shown in Fig. 4, which indicates that the behavioral data patterns were quite similar between two orders of tasks (groups). A repeated-measures ANOVA was used to analyze the interaction between advantageous selections across blocks and task order; this interaction was not significant: For the IGT, Decks × Order, $F(3, 102) = 0.20$, $ns$; for the SGT, Decks × Order, $F(3, 102) = 0.71$, $ns$. Consequently, we analyzed the data from both groups (orders) together in the following sections.

## 3. A comparison of decision learning models

The following decision learning models are based on three general assumptions. First, an individual's evaluation of the positive and negative payoffs can be represented by a unidimensional utility function. Second, expectations about payoffs for each deck are learned on the basis of the experienced utilities on each trial. Third, these expectancies determine the choice probabilities for selecting each deck on each trial. In this study, we evaluated a factorial combination of 2 Utility Functions × 2 Learning Rules × 2 Choice Probability Rules, producing a total of eight models.

### 3.1. Utility

#### 3.1.1. Expectancy utility function

The evaluation of gains and losses on trial $t$ is represented by a weighted utility function, $u(t)$:

$$u(t) = (1 - W) \cdot win(t) - W \cdot loss(t). \tag{1}$$

The terms $win(t)$ and $loss(t)$ are the amount of money won and lost on trial $t$, respectively. $W$ is the *attention to loss* parameter that denotes the weights participants place on losses over gains; the value of the parameter varies between 0 and 1 (Busemeyer & Stout, 2002). Presumably, $W$ measures reward-seeking (low $W$ value) or loss-aversive (high $W$ value) characteristics of an individual. This function implicitly assumes that decision makers process gains and losses separately according to a piecewise linear utility function.

#### 3.1.2. Prospect utility function

$$u(t) = \begin{cases} x(t)^{\alpha} & \text{if } x(t) \geq 0 \\ -\lambda |x(t)|^{\alpha} & \text{if } x(t) < 0. \end{cases} \tag{2}$$

The expectancy utility function assumes that the subjective utility is linearly proportional to the actual payoff amount. This linear approximation has been considered sufficient for small payoffs; however, this utility function cannot account for the gain–loss frequency effect. For example, getting −$1 four times seems worse than getting −$4 once for most people although the sum of losses is equivalent (Erev & Barron, 2005). However, the expectancy utility function predicts that those two events have the same overall utility. To account for the frequency effect, it may be necessary to use a nonlinear function, such as that used in prospect theory (Tversky & Kahneman, 1992).

Here, $x(t)$ is the net gain ($= win(t) - |loss(t)|$) on trial t, $\alpha$ is a parameter that governs the shape of the utility function, and $\lambda$ is a loss-aversion parameter. Unlike the expectancy utility function, this function assumes that decision makers process only the net gain, and that the net gain is nonlinearly proportional to the actual net gain. As $\alpha$ goes to zero, the shape of this function becomes more like a step function.

### 3.2. Updating of expectancies

Two general classes of models were employed: a delta learning rule and a decay-reinforcement rule.

#### 3.2.1. Delta learning rule

This rule, also known as a Rescorla–Wagner rule (Rescorla & Wagner, 1972), has been employed in connectionist theories of learning (Gluck & Bower, 1988; Rumelhart et al., 1987; Sutton & Barto, 1998), IGT studies (e.g., Busemeyer & Stout, 2002; Yechiam & Busemeyer, 2005), and binary choice studies (Yechiam & Busemeyer, 2008).

$$E_j(t) = E_j(t-1) + A \cdot \delta_j(t) \cdot [u(t) - E_j(t-1)] \tag{3}$$

Here, $E_j(t)$ refers to the expectancy for deck $j$ on trial $t$. The $A$ is the updating parameter ($0 < A < 1$), which dictates how much the expectancy of a selected deck $j$ on trial $t$ is modified by the prediction error, $[u(t) - E_j(t-1)]$. The expectancies of unselected decks remain as they were in the previous trial. The variable $\delta_j(t)$ is a dummy variable, which is 1 if deck $j$ is chosen and 0 otherwise.

#### 3.2.2. Decay-reinforcement rule

The equation of the decay-reinforcement rule (Erev & Roth, 1998) for the expectancy about deck $j$ on trial $t$ is

$$E_j(t) = A \cdot E_j(t-1) + \delta_j(t) \cdot u(t). \tag{4}$$

This rule assumes that the past expectancy is always discounted, and a selected deck on the current trial is updated by $u(t)$. $A$ is a recency parameter ($0 < A < 1$) that dictates how much the past expectancy is discounted. This rule has more flexibility than the delta rule because it permits the expectancies of all alternatives to change on each trial. Previous studies (Yechiam & Busemeyer, 2005; 2008) found that a decay-reinforcement model had a better post hoc fit than a delta learning model. However, high model flexibility may over-fit the data and lead to poor generalizability. Therefore, both rules were tested again in this study using the generalization criterion methods.

### 3.3. Choice rule

Upon first consideration, it may seem rational to always prefer the deck with the highest expected valence. This "greedy" strategy, however, leaves very little room for exploration, and the danger is that the decision maker quickly gets stuck choosing an inferior option. What is needed is some procedure to ensure that participants initially explore the decks, and only after a large number of trials decide to always prefer the deck with the highest expected valence. One of the standard reinforcement learning methods to achieve this is what is called a ratio-of-strength choice rule, which is also known as softmax selection or Boltzmann exploration (Kaelbling, Littman, & Moore, 1996; Luce, 1959). This rule assumes that decision makers use a probabilistic matching rule in which the probability of choosing an alternative is proportional to the relative strength of the alternative. Research suggests that this rule has greater accuracy for the IGT than the maximization rule with guessing (Yechiam & Busemeyer, 2005). Recall that $E_j(t)$ is the expectancy after $t$ trials of feedback. Define $D(t+1)$ as the deck chosen on the next trial, $t+1$. Then, according to the model, the probability that deck $j$ is chosen on the next trial ($Pr[D(t+1) = j]$) is determined by the following ratio rule:

$$Pr[D(t+1) = j] = \frac{e^{\theta(t) \cdot E_j(t)}}{\sum_{k=1}^{4} e^{\theta(t) \cdot E_k(t)}}. \tag{5}$$

The parameter $\theta(t)$ determines the sensitivity of the choice probabilities to the expectancies. $1/\theta(t)$ can be thought of as the "temperature" on trial $t$. When $\theta(t)$ approaches zero (the temperature is very high), choices become completely random, allowing for a lot of exploration and equivalent to random guessing. For very large values of $\theta(t)$ (at very low temperature), choices become deterministic and resemble the maximization rule in which a deck with the

highest expectancy is always selected. We examined both trial-dependent and trial-independent ratio choice rules.

*3.3.1. Trial-dependent choice rule*

Sensitivity, $\theta(t)$, may increase or decrease over trials. As participants accumulate experience as trial numbers increase, they may have more confidence in their expectancies, and $\theta(t)$ may increase accordingly. On the other hand, $\theta(t)$ may decrease as training progresses, reflecting fatigue or boredom of participants. This dynamic is formalized by the following power function:

$$\theta(t) = (t/10)^c. \tag{6}$$

Here, $c$ is a free parameter called *consistency*. Positive values of $c$ indicate increasing sensitivity (more deterministic) over time, and negative values of $c$ indicate decreasing sensitivity (more random guessing). Note that before 10 trials, the sensitivity parameter favors exploration; but after 10 trials exploitation becomes more dominant.

*3.3.2. Trial-independent choice rule*

Previous studies show that the trial-dependent choice rule explains choice behavior on the IGT and binary choice tasks quite well. However, it is unclear whether such a rule can successfully explain choices in other multi-alternative gambling tasks (e.g., SGT). Therefore, we also employed a trial-independent choice rule (Yechiam & Ert, 2007):

$$\theta(t) = 3^c - 1. \tag{7}$$

If $c$ gets close to zero (as $\theta \to 0$), the decision makers make random selections, and a large value of $c$ captures a deterministic choice rule. The consistency parameter $c$ ranges from 0 to 5 so that the sensitivity ranges from 0 (random) to 242 (almost deterministic).

## 4. Model evaluation

Three different model evaluation methods were used to compare the 8 models. One method was to compare the post hoc fits of each model to each task, and the other two methods were based on the generalization criterion. The details of each model comparison method and the results of each method are described in the following.

### *4.1. Model fit using the one-step-ahead prediction method*

We used the one-step-ahead prediction method to compute model fits for each person and task. This method examines the accuracy of a model in predicting an individual's choice on the next trial based on an individual's actual sequence of choices and consequent payoffs up to and including the current trial. Define $Y_i(t)$ as the sequence of choices made by an individual $i$ up to and including trial $t$, and define $X_i(t)$ as a vector with a corresponding sequence of payoffs produced by these choices. The model generates the probability that the individual will select an alternative $D_j$ ($j$ is from 1 to 4) on the next trial $t + 1$, which is

$Pr[D_j(t+1)|X_i(t), Y_i(t)]$. The accuracy of model predictions is measured by log likelihood (LL) criterion for each individual:

$$\begin{aligned} \mathrm{LL}_{\mathrm{Model},\, i} &= \ln Pr(\text{Data for person } i \,|\, \text{Model}) \\ &= \sum_{t=1}^{n-1} \sum_{j=1}^{4} \ln \left( Pr[D_j(t+1) \,|\, X_i(t), Y_i(t)] \right) \cdot \delta_j(t+1). \end{aligned} \tag{8}$$

Here, $n$ is the number of trials (= 120) and $\delta_j(t+1)$ is a dummy variable that is 1 if deck $j$ is chosen on trial $t+1$ and 0 otherwise. Each model has three ($W$, $A$, and $c$) or four ($\alpha$, $\lambda$, $A$, and $c$) free parameters, and the parameter values were estimated using the maximum likelihood estimation. A computer search algorithm was used to search for the set of parameters that optimized the LL, which was based on a combination of grid-search and simplex search methods (Nelder & Mead, 1965): The algorithm used 50 different starting positions on a grid and found the parameters that maximized the LL for each starting position. Then we selected the solutions of a starting point that produced the global maximum across all grid points.[4]

The grid for the parameter search used the following constraints: $W$, $\alpha$, and $A$ were limited to values between 0 and 1; $\lambda$ was limited to values between 0 and 10; trial-dependent consistency was limited to values between $-5$ and 5; and trial-independent consistency was limited to values between 0 and 5.

To calculate the measure of fit, each model was compared with a baseline statistical model by computing differences in LL as follows:

$$G^2 = 2 \cdot [\mathrm{LL}_{\mathrm{model}} - \mathrm{LL}_{\mathrm{baseline}}]. \tag{9}$$

The Bernoulli baseline model selects an alternative using a statistical Bernoulli process: Probabilities of each alternative are assumed to be constant for all trials (equal to the individual's overall proportion of each deck) so the choices are statistically independent across trials. Therefore, a cognitive model can outperform the Bernoulli baseline model only when it explains trial-to-trial dependence or learning effects. Positive values of $G^2$ indicate that the model explains the data better than the Bernoulli baseline model.

A particular model may have a better fit simply because it has more parameters. To consider model complexity and adjust for differences in the number of parameters, we used the Bayesian Information Criterion (BIC; Schwartz, 1978):

$$\mathrm{BIC} = G^2 - \Delta k \cdot \ln(N). \tag{10}$$

$N$ is the number of observations (or trials) and $\Delta k$ is the difference in the number of parameters between a cognitive model and the baseline model. Here, a positive BIC would indicate that the model is better than the baseline model even after model complexity is considered. Equation 10 shows that the $G^2$ value of models with four parameters is reduced by $\Delta k \cdot \ln(N) = 1 \cdot \ln(120) \approx 4.79$ compared to the $G^2$ of models with three parameters.

Tables 3 and 4 present statistics that summarize the distribution of parameters obtained from the 36 participants for each of the eight models (Table 3 for the IGT and Table 4 for the SGT). The distributions are summarized in terms of the means and standard deviations of

Table 3
The means and standard deviations (in parentheses) of Bayesian Information Criterion (BIC) scores and estimated parameters of the models in the Iowa Gambling Task

| Utility | Updating | Choice | BIC | $A$ | W or λ | $c$ | $\alpha$ |
|---|---|---|---|---|---|---|---|
| EU | DEL | TDC | 7.45 (32.2) | 0.14 (0.2) | 0.42 (0.3) | 1.49 (1.6) | na |
| | | TIC | 7.49 (25.4) | 0.13 (0.3) | 0.40 (0.2) | 3.16 (1.7) | |
| | DRI | TDC | 18.25 (35.4) | 0.67 (0.4) | 0.41 (0.3) | 0.19 (1.4) | |
| | | TIC | 17.20 (33.2) | 0.68 (0.4) | 0.36 (0.3) | 0.81 (0.8) | |
| PU | DEL | TDC | 12.98 (30.8) | 0.17 (0.3) | 2.75 (3.8) | 1.12 (1.3) | 0.53 (0.4) |
| | | TIC | 11.17 (24.7) | 0.17 (0.3) | 2.56 (3.4) | 2.81 (1.8) | 0.55 (0.4) |
| | DRI | TDC | 7.12 (37.0) | 0.67 (0.3) | 1.47 (2.4) | −0.24 (0.5) | 0.47 (0.4) |
| | | TIC | 15.48 (31.3) | 0.75 (0.3) | 2.19 (3.4) | 0.41 (0.4) | 0.32 (0.4) |

*Note*. EU = expectancy utility; PU = prospect utility; DEL = delta learning rule; DRI = decay-reinforcement learning rule; TDC = trial-dependent consistency; TIC = trial-independent consistency.

the parameters. In the IGT, every model had positive average BICs relative to the Bernoulli baseline model (every model was better than the baseline model). In the SGT, all models with the expectancy utility function had negative BICs except one, whereas models with the prospect utility function still all had positive BICs. Regarding learning rules, overall the decay-reinforcement rule had higher accuracy (higher BIC scores), which is consistent with previous research (Yechiam & Busemeyer, 2005; 2008).

Next, we examined the model fits using a random baseline model. The random baseline model assumes that each choice option is equally likely across all trials. This baseline model is important because it can be applied to the generalization test phase. The Bernoulli baseline model cannot be used for generalization tests because it requires estimation of new parameters. Using the parameter estimates of each individual from each task, models made one-step-ahead predictions of the same task. A detailed procedure is described in Appendix A. The shaded or gray regions of Table 5 presents statistics that summarize the distribution of $G^2$ statistics

Table 4
The means and standard deviations (in parentheses) of Bayesian Information Criterion (BIC) scores and estimated parameters of the models in the Soochow Gambling Task

| Utility | Updating | Choice | BIC | $A$ | W or λ | $c$ | $\alpha$ |
|---|---|---|---|---|---|---|---|
| EU | DEL | TDC | −6.78 (20.6) | 0.31 (0.3) | 0.31 (0.4) | −0.31 (1.6) | na |
| | | TIC | −6.61 (18.8) | 0.22 (0.3) | 0.26 (0.4) | 2.05 (1.8) | |
| | DRI | TDC | −0.62 (32.2) | 0.66 (0.3) | 0.39 (0.4) | −0.81 (2.0) | |
| | | TIC | 1.57 (27.3) | 0.75 (0.3) | 0.24 (0.4) | 0.51 (0.7) | |
| PU | DEL | TDC | 1.23 (20.3) | 0.24 (0.3) | 2.72 (3.6) | −0.09 (1.1) | 0.45 (0.4) |
| | | TIC | 2.30 (19.3) | 0.24 (0.3) | 2.99 (4.2) | 1.97 (1.8) | 0.24 (0.3) |
| | DRI | TDC | 4.28 (32.1) | 0.62 (0.4) | 1.47 (2.9) | −0.39 (0.5) | 0.30 (0.3) |
| | | TIC | 10.73 (28.7) | 0.71 (0.3) | 1.96 (3.4) | 0.37 (0.3) | 0.36 (0.4) |

*Note*. EU = expectancy utility; PU = prospect utility; DEL = delta learning rule; DRI = decay-reinforcement learning rule; TDC = trial-dependent consistency; TIC = trial-independent consistency.

Table 5
The distributions of $G^2$ scores

| | Model | | | Estimated Task: IGT | | | | | Estimated Task: SGT | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Target task | Utility | Updating | Choice | L10% | M | Mdn | U90% | % | L10% | M | Mdn | U90% | % |
| IGT | EU | DEL | TDC | 4.2 | 68.5 | 51.7 | 155.7 | 94.4 | −51.2 | 1.6 | 0.2 | 53.5 | 50.0 |
| | | | TIC | 7.4 | 71.1 | 43.1 | 164.1 | 100.0 | −40.2 | 12.6 | 6.0 | 64.2 | 58.3 |
| | | DRI | TDC | 14.2 | 80.8 | 73.8 | 154.6 | 100.0 | −35.1 | 0.0 | 7.5 | 80.3 | 63.9 |
| | | | TIC | 16.6 | 80.3 | 70.6 | 155.4 | 100.0 | −4.9 | 21.8 | 20.3 | 119.4 | 86.1 |
| | PU | DEL | TDC | 10.0 | 80.3 | 66.2 | 154.7 | 100.0 | −12.5 | 19.6 | 11.2 | 71.2 | 66.7 |
| | | | TIC | 9.2 | 78.5 | 60.4 | 177.6 | 100.0 | −9.9 | 23.1 | 20.7 | 81.8 | 77.8 |
| | | DRI | TDC | −3.3 | 75.0 | 66.3 | 158.5 | 88.9 | −19.5 | 33.3 | 20.0 | 108.1 | 80.6 |
| | | | TIC | 17.1 | 83.4 | 80.5 | 154.0 | 100.0 | −23.1 | 36.7 | 31.2 | 118.7 | 77.8 |
| SGT | EU | DEL | TDC | −509.9 | −131.0 | −29.6 | 38.2 | 25.0 | 2.9 | 37.4 | 22.3 | 115.8 | 100.0 |
| | | | TIC | −277.5 | −78.3 | −29.2 | 52.7 | 30.6 | 2.7 | 38.6 | 23.7 | 114.2 | 97.2 |
| | | DRI | TDC | −348.4 | −86.0 | −41.6 | 68.6 | 38.9 | 1.6 | 44.4 | 28.6 | 112.8 | 94.4 |
| | | | TIC | −222.2 | −41.6 | −17.3 | 68.8 | 41.7 | 3.3 | 46.7 | 33.6 | 110.7 | 97.2 |
| | PU | DEL | TDC | −441.5 | −105.7 | −20.0 | 87.6 | 41.7 | 8.1 | 50.8 | 31.7 | 129.4 | 100.0 |
| | | | TIC | −273.4 | −64.3 | −15.6 | 91.5 | 44.4 | 7.2 | 51.7 | 29.0 | 136.2 | 100.0 |
| | | DRI | TDC | −344.6 | −62.7 | −6.8 | 81.4 | 47.2 | 2.1 | 54.2 | 37.1 | 129.9 | 94.4 |
| | | | TIC | −230.6 | −37.1 | −4.2 | 91.2 | 41.7 | 14.4 | 60.6 | 39.8 | 130.1 | 100.0 |

*Note*. Table shows lower 10% bounds, means, medians, upper 10% bounds, and percentage of individuals for which the generalization prediction of a decision learning model is better than that of a random model. Shaded cells denote tests that do not involve generalization. EU = expectancy utility; PU = prospect utility; DEL = delta learning rule; DRI = decay-reinforcement learning rule; TDC = trial-dependent consistency; TIC = trial-independent consistency; IGT = Iowa Gambling Task; SGT = Soochow Gambling Task; M = Mean; Mdn = Median.

obtained from the 36 participants for each of the eight models. $G^2$ are used here rather than BICs because we later compare these results with those obtained from a generalization test, and the BIC is not relevant for these tests. In particular, the shaded region in the upper left-hand corner under the column labeled "Estimated Task: IGT" and the row labeled "Target task: IGT" presents the $G^2$ results for the fits to the IGT; the shaded region in the lower right-hand corner under the column labeled "Estimated Task: SGT" and the row labeled "Target task: SGT" presents the $G^2$ results for the fits to the SGT. The distributions are summarized in terms of the lower 10th percentile, mean, median, and 90th percentile of the $G^2$ scores. A positive $G^2$ value for a model would imply that the model can make better one-step-ahead predictions than a random model, on average. Also, a percentage greater than 50% would imply that the model is a better fit than a random model for more than 50% of participants, which is a chance level.

Generally, the $G^2$ results in the shaded regions of Table 5 show that the prospect utility function performed better than the expectancy utility function. Given that a prospect utility function is used, the decay-reinforcement learning model performed slightly better than the delta learning model. Finally, the prospect utility function–decay reinforcement learning–trial independent choice model performed best.

### *4.2. Generalization at the individual level*

Using the generalization criterion method, we tested generalizability of the models across tasks at the individual level (Yechiam & Busemeyer, 2008). Parameter values estimated from one task (either the IGT or the SGT) were used to make one-step-ahead predictions again using the procedure in Appendix A, but this time predictions were made for the other task. $G^2$ and the percentage of individuals for which $G^2 > 0$ were used as fit indexes. Again, the baseline model was a random model (equal chance of selecting each deck).

The unshaded, or white regions, in Table 5 show the results of the one-step-ahead generalization predictions. In particular, the region in the upper right-hand corner shows the results when the parameters estimated from the SGT were used to predict choices in the IGT; the region in the lower left corner shows the results when the parameters from the IGT were used to predict choices in the SGT. Generally, the same results were found with this generalization test as found before with the model fitting method. The prospect utility function performed better than the expectancy utility function; given that a prospect utility function is used, the decay-reinforcement learning model performed better than the delta learning model; finally, the prospect utility function decay-reinforcement learning trial-independent choice model performed best, although the trial-dependent choice model had a higher percentage of positive $G^2$.

The results also show that the generalization from the SGT to the IGT was more successful than generalization from the IGT to the SGT. When the SGT was used to predict the IGT, all models had positive average $G^2$ values and were above or equal to chance level percentage (50%). Specifically, models with the prospect utility function had close to 80% positive $G^2$ values, on average. In contrast, when the IGT was used to predict the SGT, all models had negative average $G^2$ scores and less than 50% accuracy compared to a random model.

This is consistent with the result of Yechiam and Busemeyer (2008), in which a task (high-variance task) with the poorest BIC fit and the poorest learning nevertheless led to the best generalizability of one-step-ahead predictions. We discuss this issue further in the General Discussion.

### *4.3. Simulation method*

The one-step-ahead prediction method uses the observed past history of choices for an individual on a task to predict the next choice. An alternative method, called the simulation method (Yechiam & Busemeyer, 2005), uses the parameters estimated from one task to make a priori predictions for the entire sequence of 120 choices in the second task (without using any information about observed choices on the second task). Strictly, it is not an individual level analysis. However, unlike previous studies (e.g., Yechiam & Busemeyer, 2005), we ran simulations for each participant separately and averaged all simulation results afterward.

For each participant, the previously described parameter estimates of one task (estimated using the one-step-ahead prediction method) were used to generate 100 simulations for the entire sequence of choices for both tasks, producing 3,600 simulations in total (100 × 36 participants = 3,600). The 100 simulations generated from each participant were averaged across all participants to produce the predicted probability of choice on each trial for each task, and these predicted probabilities were then compared to the observed choice probabilities using mean square deviation (MSD) scores[5] (a detailed procedure for the simulation method is in the Appendix B):

$$\mathrm{MSD} = \frac{1}{4 \cdot n} \sum_{t=1}^{n} \sum_{j=1}^{4} (\bar{D}_{\mathrm{exp},j}(t) - \bar{D}_{\mathrm{sim},j}(t))^2. \tag{11}$$

Here, $n$ is the total number of trials (= 120), $\bar{D}_{\mathrm{exp},j}(t)$ is the mean percentage of choices from alternative $j$ on trial $t$ across all participants' experimental data, and $\bar{D}_{\mathrm{sim},j}(t)$ is the mean percentage of choices from alternative $j$ on trial $t$ across all participants' simulation data. When calculating MSDs, we used percentage scores (e.g., 35%) instead of proportions (e.g., 0.35). The MSD is a measure of deviation of simulation data from the observed data, so the lower the MSD is, the higher its accuracy. Before computing the MSD scores, we smoothed experimental data by running a moving window of seven trials (each datum point is the average of nearby 7 data).

Table 6 summarizes the results of simulation data. The shaded regions (upper left corner and lower right corner) show the results when the parameters from a task are used to generate the simulated choices for the same task (post hoc fits). The unshaded or white regions (upper right corner or lower left corner) show the results when the parameters from one task are used to generate simulated choices for a different task (a priori predictions).

The results in Table 6 clearly show that, as found before, models with the prospect utility function outperformed models with the expectancy utility function. However, unlike before, generally models with the delta learning rule performed better than models with the decay-reinforcement learning rule. With regard to the consistency models, neither model seemed to outperform the other. Regarding performance of learning rules, Yechiam and Busemeyer

Table 6
MSDs of each model between behavioral data and model predictions

| | Model | | | Estimated Task: IGT | Estimated Task: SGT |
|---|---|---|---|---|---|
| Target task | Utility | Updating | Choice | MSD | MSD |
| IGT | EU | DEL | TDC | 80.71 | 153.40 |
| | | | TIC | 83.80 | 219.09 |
| | | DRI | TDC | 120.19 | 204.61 |
| | | | TIC | 166.17 | 291.43 |
| | PU | DEL | TDC | 24.58 | 90.79 |
| | | | TIC | 24.12 | 84.43 |
| | | DRI | TDC | 91.60 | 161.67 |
| | | | TIC | 87.85 | 151.61 |
| SGT | EU | DEL | TDC | 222.33 | 77.87 |
| | | | TIC | 272.83 | 109.47 |
| | | DRI | TDC | 115.72 | 65.09 |
| | | | TIC | 151.77 | 146.38 |
| | PU | DEL | TDC | 25.95 | 23.61 |
| | | | TIC | 23.12 | 21.63 |
| | | DRI | TDC | 22.27 | 25.37 |
| | | | TIC | 35.08 | 26.27 |

*Note*. Shaded cells denote tests that do not involve generalization. EU = expectancy utility; PU = prospect utility; DEL = delta learning rule; DRI = decay-reinforcement learning rule; TDC = trial-dependent consistency; TIC = trial-independent consistency; MSD = mean square deviation; IGT = Iowa Gambling Task; SGT = Soochow Gambling Task.

(2005, partial information condition) found similar results in a study that only used the IGT: The decay-reinforcement learning model was clearly superior with one-step-ahead predictions; but this superiority largely disappeared when the entire sequence of choices was simulated without using past choices.

Fig. 5 visualizes the simulation results of the prospect utility–delta learning–trial independent choice model, one of the best among all eight models. As can be seen in this figure, this model's performance was excellent in post hoc fits (IGT→IGT and SGT→SGT). The MSD of this model was below 25, meaning that the average discrepancy between the simulation data and the observed data was less than 5% on each trial. It not only predicted the preference of Deck B over Deck A on the IGT, but also predicted more choices from disadvantageous decks on the SGT, whereas the models with the expectancy utility function could explain neither. For conciseness, only the simulation results for the prospect utility are shown.

In generalization tests, this model with the delta learning rule showed acceptable performance for IGT→SGT, which is consistent with the results of Table 6. Meanwhile, Fig. 6 shows that simulation results of the model with the decay-reinforcement learning rule (with the prospect utility and trial-independent choice model) are worse than those of the previous model. Even in the IGT, this model predicted too many choices from Deck B and too few choices from Deck C. In a generalization test from IGT→SGT, its predictions are qualitatively correct but not as good as the predictions of the delta learning rule.

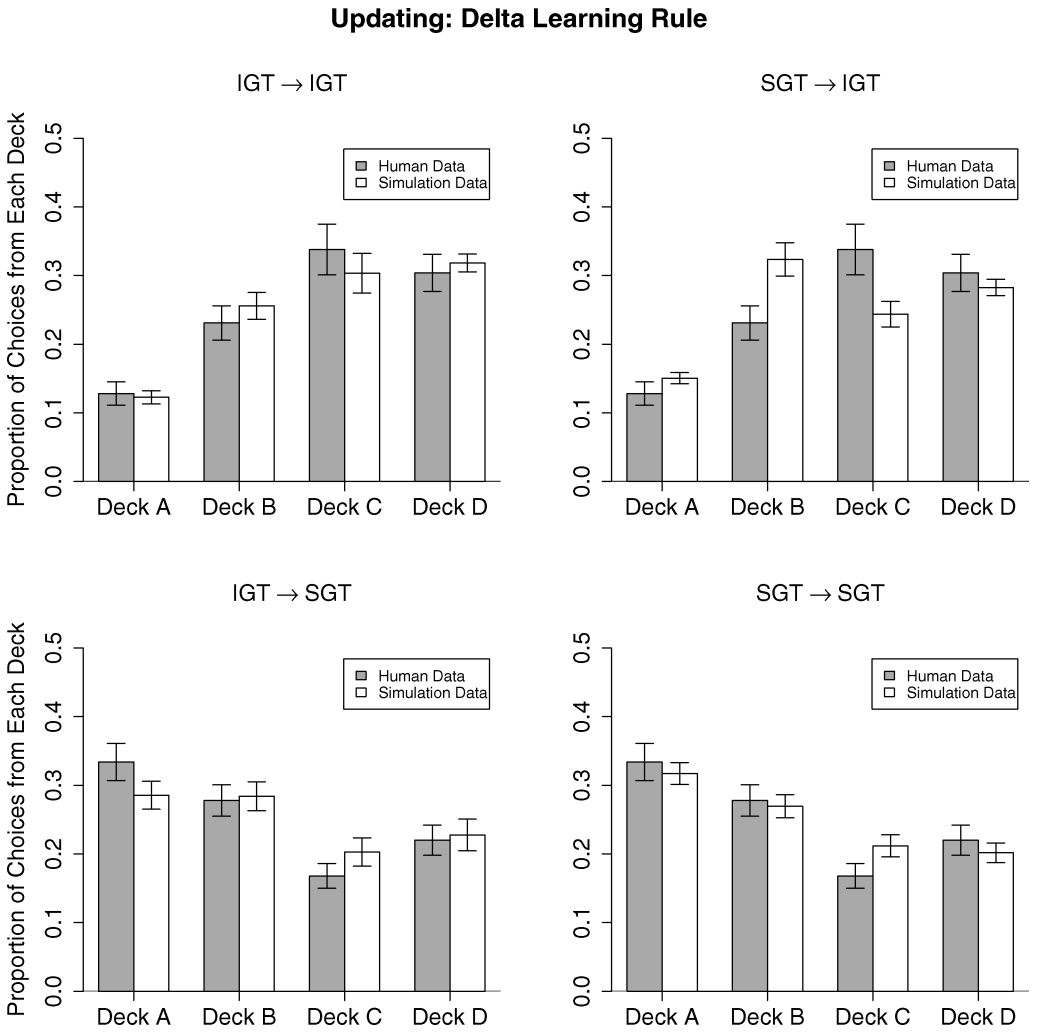


Fig. 5. Simulation results of the prospect utility function delta learning rule trial-independent model: calibration set → test set. Error bars indicate ±1 SEM. IGT = Iowa Gambling Task; SGT = Soochow Gambling Task.

### *4.4. Resolving the difference between model comparison methods*

Why did the delta learning rule perform better than the decay-reinforcement learning rule in simulation results? The clearest difference between the one-step-ahead prediction method and the simulation of all choices is that the former makes use of the actual past choices to predict the next choice, whereas the latter does not use any information about past choices to make predictions. The decay-reinforcement learning model can track past behavior (because it decays unchosen options toward zero), which may give it an advantage when it needs to use past behavior to predict the next choice. This advantage is removed with the simulation method; and, once removed, the delta learning rule outperforms the decay-reinforcement learning rule.

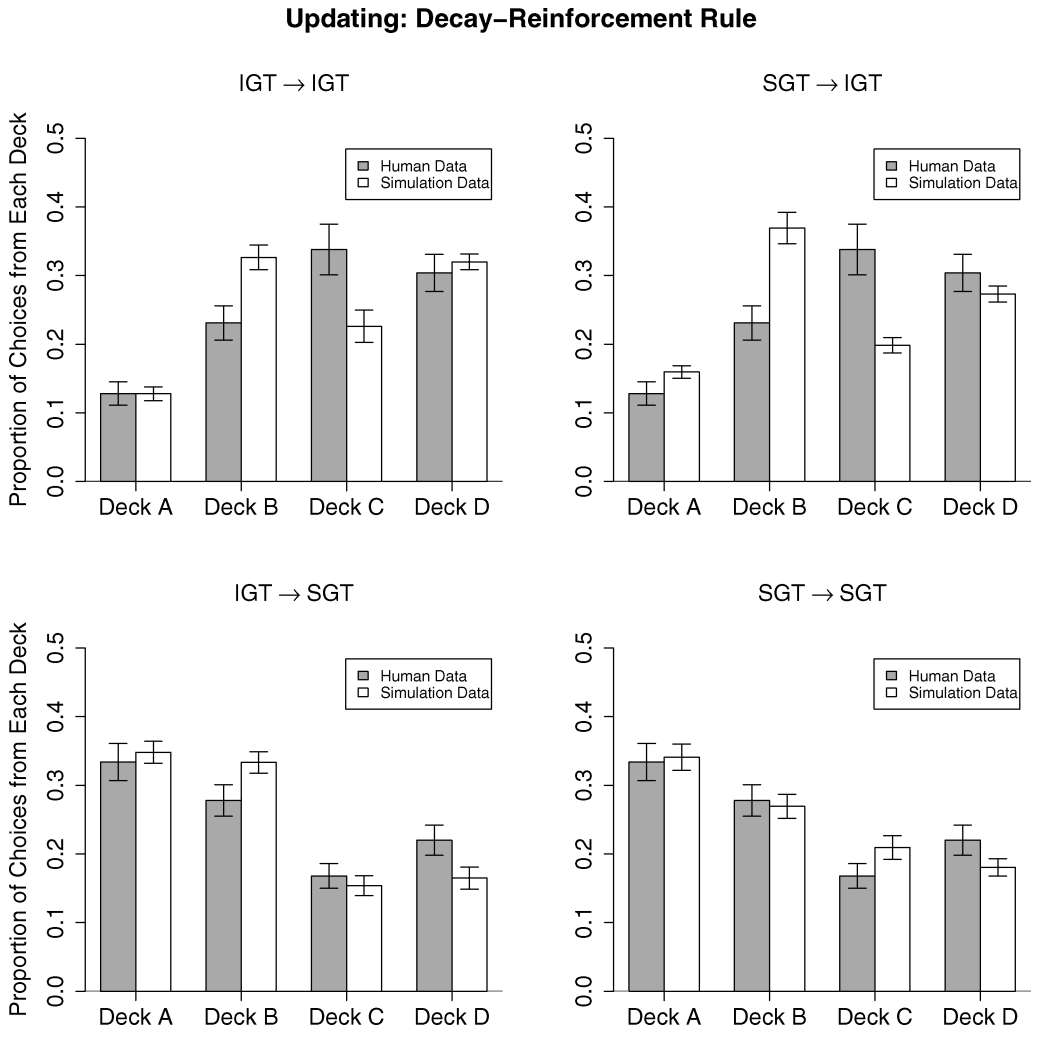


Fig. 6. Simulation results of the prospect utility function decay-reinforcement learning rule trial-independent model: calibration set → test set. Error bars indicate ±1 SEM. IGT = Iowa Gambling Task; SGT = Soochow Gambling Task.

Yechiam and Ert (2007) evaluated the degree of reliance of model predictions on past choices and past payoffs. They found that the delta learning rule relied more on past payoffs than the decay-reinforcement learning rule, and it gave an advantage to the delta learning rule when making predictions to new conditions. In making predictions about new conditions, which presumably have new payoff distributions, a model needs to base its predictions on payoffs rather than choices.

This also seems to explain why the simulation results from the IGT to the SGT were better than from the SGT to the IGT. Note that one-step-ahead predictions from the SGT to the IGT were better than the IGT to the SGT. These findings may appear to be contradictory or

inconsistent but, in fact, they are not. The IGT has more payoff variability (or entropy) than the SGT (Tables 1 and 2) implying that parameters estimated from the IGT, compared to the SGT, were based more on past payoffs rather than past choices. Therefore, simulation predictions using parameters from the IGT would be more generalizable than using parameters from the SGT to new conditions. As mentioned earlier, the one-step-ahead prediction method relies heavily on past choices. Therefore, when making one-step-ahead predictions, predictions that are based on past choices would have more advantages. We suspect that is why the one-step-ahead predictions from the SGT to the IGT are better than the other way around. As shown in Fig. 3, choices on the SGT were more distributed across decks than the IGT and perhaps models could get better parameter estimates from the SGT data for one-step-ahead predictions.

## 5. General discussion

The purpose of this article was to compare several models of learning in two decision-making tasks using the generalization criterion methods. A new experiment was reported that investigated learning processes in two different types of gambling tasks—the IGT and the SGT. The two tasks produced dramatically different learning curves: The former showed much greater learning and a different direction of learning compared to the latter. These empirical results provided the data sets for evaluating the decision learning models.

Eight models were compared by examining two utility assumptions (expectancy utility vs. prospect utility) crossed with two learning assumptions (delta learning vs. decay-reinforcement learning) crossed with two choice rules (trial-dependent vs. trial-independent choice consistency). This design for model comparison allows testing of each property of the model independently of other model properties (Yechiam & Busemeyer, 2005).

We used three different methods to evaluate the models at the individual level of analysis. The first method used a post hoc fit criterion, which evaluated the accuracy of one-step-ahead predictions after fitting the parameters to both gambling tasks. The second method used a generalization criterion, which evaluated the accuracy of one-step-ahead predictions for one of the gambling tasks using parameters estimated from the other gambling task. The third method also used a generalization criterion, but the evaluation was based on the accuracy for simulating all of the choices in a second task using parameters estimated from a first (calibration) task.

The results obtained by using one-step-ahead predictions turned out to be very similar for model fits and generalization tests. Both methods found that (a) the prospect utility function performed much better than the expectancy utility function; (b) given the prospect utility function, the decay-reinforcement learning model moderately performed better than the delta learning model; and (c) given the prospect utility function and the decays reinforcement learning rule, the trial-independent choice rule performed slightly better than the trial-dependent choice rule. These results were generally consistent across both the IGT and the SGT. Model fit indexes, as well as generalization test indexes, were better for the IGT than the SGT. Apparently, it was more difficult to predict behavior in the SGT using the one-step-ahead prediction method. This may reflect the greater difficulty in learning that occurs in the SGT or better parameter estimates for the one-step-ahead predictions from the SGT to the IGT as mentioned earlier.

The third method (based on simulating all of the choices) produced different results than the one-step-ahead predictions. The prospect utility function still performed better than the expectancy utility function, and there was little or no difference between trial-independent and trial-dependent choice rules. However, in this case the delta learning rule outperformed the decay-reinforcement learning rule. This is believed to be due to models' different reliance on past payoffs and past choices. The delta learning rule depends more on past payoffs than the decay-reinforcement learning rule (Yechiam & Ert, 2007). Therefore, its performance is worse in making one-step-ahead predictions, which uses observed choices but it has better generalizability and better performance on the simulation results.

Then, a critical question is, "What is the best model, or which model should be used in making predictions?" Some readers might get an impression that the choice of a learning model depends on which measure of fit is used. Of course, this is not true and it is not the conclusion of this article. We argue that the choice of "the best" model often depends on the types (or goals) of predictions. As Erev and Haruvy (2005) pointed out, the best model in predicting tomorrow's weather can be different from the best model in predicting the next year's weather (e.g., Vintzileos, Delecluse, & Sadourny, 1999), so one's best model for weather forecasting depends on his or her need. All of the models considered here are quite simple; and, for this reason, they are all imperfect. Nevertheless, they may be useful for prediction under limited circumstances. In this study, regarding learning rules, the decay-reinforcement learning rule was better in making prediction for the very next trial (short-term prediction), whereas the delta learning rule was better in making predictions for the entire sequence of choices (long-term prediction). As an example, for the purpose of "adaptive testing" (Erev & Haruvy, 2005) or on-the-fly manipulation on the task, the decay-reinforcement learning rule would be better. In such a situation, a model needs to predict participants' next responses as accurately as possible based on past choices to change or manipulate payoffs of decks, and the decay-reinforcement learning model is a better choice according to the results of this study. On the other hand, to predict a participant's or a group's performance on a new gambling task, the delta learning rule should be used. For instance, one may want to develop a better gambling task than the IGT to measure decision-making deficits of substance abusers. Then, a new task or payoff distributions can be tested using parameters of substance abusers estimated from the IGT, as long as the model's performance in new conditions proved to be adequate. Previous researchers (e.g., Erev & Haruvy, 2005; Yechiam & Busemeyer, 2005) found that different models emerge as best depending on the chosen objective (short term vs. long term). Our results are consistent with Yechiam and Busemeyer (2005), and it is a very consistent finding across other components of models and target tasks. Our model design (the 2 Utility Functions $\times$ 2 Learning Rules $\times$ 2 Choice Rules of 8 models) allowed us to rigorously test the short-term and long-term predictability of each learning model. As Tables 5 and 6 show, the previous finding was generally true whether the utility function was either the expectancy utility or the prospect utility, whether the choice rule was trial-dependent or trial-independent, or whether the target task was the IGT or the SGT.

Regarding utility functions, the prospect utility function outperformed the expectancy utility in almost every measure. Numerous previous studies on prospect theory exist but they were mostly on "decisions from description" studies. This study suggests that its utility function accounts for at least some data in experience-based tasks as well. The success of the prospect

utility function lies in its nonlinear shape (diminishing sensitivity). Because large payoffs are discounted more than small payoffs, it can account for the gain–loss frequency effect. This is important in explaining decisions made on the IGT because most of learning toward advantageous decks of healthy participants comes from fewer choices from Deck B, which has the net loss frequency of 1 out of 10. At the same time, drug abusers and many other clinical populations fail to avoid choosing from Deck B, which is the main reason why they show poor performance on the IGT. Recent studies by Desmeules, Bechara, and Dube (2007) and Erev, Ert, and Yechiam (in press) also found that diminishing sensitivity plays an important role in decision making.

In this article, the generalization criterion method was applied to relatively well-behaved mathematical models with just three or four free parameters. The same approach can be applied to more complex models such as neural network models or computational models based on a cognitive architecture. The generalization criterion automatically considers model complexity in making a priori predictions. Therefore, if a model is complex enough but not general enough, its fit to a calibration set will be good but its a priori predictions to new conditions will not be. In addition, the generalization method is not restricted to make a priori predictions for behavior only. For example, in Anderson and Qin (2008), an ACT–R (Anderson et al., 2004) model was fit to behavioral data, and it generated a priori predictions for the functional magnetic resonance imaging signal in several brain regions to test the model's generalizability. Brown & Braver (2005) compared two computational models (error-likelihood model and conflict model) of anterior cingulate cortex, and Hampton, Bossaerts, & O'Doherty (2006) compared a reinforcement learning model and a Markov model with higher order structure using similar frameworks. The underlying idea of all those studies is the same: estimating parameters from one measure and making a priori predictions to another measure to test a model's generalizability.

In closing, three general lessons were learned from this study. First, using the one-step-ahead prediction method, comparisons of model fits based on the BIC index very closely matched the comparisons of model accuracy based on the generalization criterion. Second, regarding basic learning processes in simple gambling tasks, the learning model selected as "the best model" depended on the objectives that we wish to achieve. If we wish to achieve short-term prediction about what happens next given observations about the past history, then the decay-reinforcement learning model is best; but if we wish to achieve long-term prediction without knowing anything about the past history of choices, then the delta learning rule is best. Third, the prospect utility function had better accuracy and generalizability than the expectancy utility function. In some of our previous work aimed at disentangling processes underlying gambling tasks, we used the Expectancy Valence Learning model (Busemeyer & Stout, 2002), which has the expectancy utility function. However, the results of the present study suggest that we should consider switching to the prospect utility function in the future.

## Acknowledgments

This research was supported by Grant R01-DA014119 from the National Institute on Drug Abuse. We thank Anthony J. Bishara and Timothy J. Pleskac for helpful discussions;

Sarah Queller, Eric Dimperio, Ryan K. Jessup, and three anonymous reviewers for helpful comments on a draft; and Bethany Ward-Bluhm and Adolfo R. Blanco for their assistance collecting data.

## Notes

1. Regarding bonuses, participants received one half of whatever money they earned at the end of each task (e.g., if a participant ended up with $20 in the Iowa Gambling Task, she or he received $10 for the task). They started each task with $20.00 and were informed about the bonus rate. If they ended up with a negative amount of play money on a task, their bonus for the task was $0.
2. One may wonder why in later sections learning models are applied to the Soochow Gambling Task (SGT) in which participants did not show "learning" if learning is defined as to choose more cards from advantageous decks as trials proceed. Note that models compared in this study are not just learning models but decision learning models. Even if the proportion of choices from advantageous decks is below 50% on average, it does not necessarily imply that there is no trial-to-trial dependence or participants made choices randomly. In fact, participants in the SGT selected significantly more cards from disadvantageous decks, on average, save Blocks 5 and 6, and it may indicate that they did not make random choices. A good decision learning model should be able to capture such a pattern as well. This is especially important to successfully model a clinical population that usually fails to choose more cards from advantageous decks even in the Iowa Gambling Task.
3. In the Soochow Gambling Task, disadvantageous decks (Decks A and B) have the equal probability of net negative payoffs. As expected, neither deck was significantly preferred over the other, $t(35) = 1.36$, $p = .18$. This was also true for Decks C and D, $t(35) = -1.87$, $p = .07$.
4. The sum of one-step-ahead prediction errors can be used for multiple purposes. In this work, we use it for *parameter estimation*; that is, we adjust the model parameters until we find the point estimates that minimize the sum of one-step-ahead prediction errors. It is important to realize that this sum of prediction errors is always fully available to the minimization algorithm. This means that the best parameter values may be those that yield relatively large prediction errors early in the series of observations, but compensate for their initial poor predictions by producing outstanding predictions at a later stage. In other work, the sum of one-step-ahead prediction errors has been used for *model selection* (e.g., Dawid, 1984; Shiffrin, Lee, Wagenmakers, & Kim, in press; Wagenmakers, Grünwald, & Steyvers, 2006). The crucial difference is that for model selection, the algorithm does not take into account the prediction errors that may occur in the future. This is similar to how meteorologists predict the weather—future observations are unavailable, and parameters cannot be adjusted based on prediction errors that have yet to occur.
5. Mean square deviation (MSD) scores were used instead of $G^2$ for two reasons. First, exact calculation of the likelihood for data from an individual requires computing the

probabilities for all possible choice sequences from the model for 120 trials to evaluate the likelihood of the observed sequence. It is incorrect to simply use the product of the probabilities for choices across trials because independence does not hold. Unfortunately, it is too difficult to compute the exact likelihood for all sequences from the model. Second, MSD scores are more intuitive for examining how good a model is in explaining overall choice patterns. For instance, a MSD score of 9, 16, or 25 stands for 3%, 4%, or 5% average discrepancy of simulations and the observed data.

## References


Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review, 111*, 1036–1060.

Anderson, J. R., & Qin, Y. (2008). Using brain imaging to extract the structure of complex events at the rational time band. *Journal of Cognitive Neuroscience, 20*, 1624–1636.

Barren, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making, 16*, 215–233.

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

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275*, 1293–1295.

Berry, D. A., & Fristedt, B. (1985). *Bandit problems: Sequential allocation of experiments*. London: Chapman & Hall.

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science, 307*, 1118–1121.

Browne, M. (2000). Cross-validation methods. *Journal of Mathematical Psychology, 44*, 108–132.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara Gambling Task. *Psychological Assessment, 14(3)*, 253–262.

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

Chiu, Y. C., Lin, C. H., Huang, J. T., Lin, S., & Huang, J. T. (2006, September). *Reexamining the effect of long-term outcome and gain-loss frequency: From uncertainty to certainty.* Paper presented at the 4th annual meeting of the Society for Neuroeconoimics, Park City, Utah.

Chiu, Y. C., Lin, C. H., Huang, J. T., Lin, S., Lee, P. L., & Hsieh, J.-C. (2008). Immediate gain is long-term loss: Are there foresighted decision makers in the Iowa Gambling Task? *Behavioral and Brain Functions, 4*, 13.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.

Dawid, A. P. (1984). Statistical theory: The prequential approach. *Journal of the Royal Statistical Society A, 147*, 278–292.

Desmeules, R., Bechara, A., & Dube, L. (2007). Subjective valuation and aymmetrical motivational systems: Implications of scope insensitivity for decision making. *Journal of Behavioral Decision Making, 21(2)*, 211–224.

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review, 112(4)*, 912–931.

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

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

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

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review, 57*, 94–107.

Forster, M. R. (2000). Key concepts in model selection: Performance and generalizability. *Journal of Mathematical Psychology, 44*, 205–231.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*, 227–247.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience, 26*, 8360–8367.

Hertwig, R., Barren, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*, 534–539.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research, 4*, 237–285.

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

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

Rescorla, R. A., & Wagner, A. R. (1972). *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement*. New York: Appleton-Century-Crofts.

Rieskamp, J., Busemeyer, J. R., & 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.

Rumelhart, D. E., McClelland, J. L., & The PDP Research Group. (1987). *Parallel distributed processing: Explorations in the micro structure of cognition* (Vols. 1 & 2). Cambridge, MA: MIT Press.

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

Shiffrin, R. M., Lee, M., Wagenmakers, E.-J., & Kim, W. (in press). *Model evaluation and selection: Established methods and recent development*.

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

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulaive representations of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.

Vintzileos, A., Delecluse, P., & Sadourny, R. (1999). On the mechanisms in a tropical ocean-global atmosphere coupled general circulation model. Part I: Mean state and the seasonal cycle. *Climate Dynamics, 15(1)*, 43–62.

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

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., & Busemeyer, J. R. (2008). Evaluating generalizability and parameter consistency in learning models. *Games and Economic Behavior, 63*, 370–394.

Yechiam, E., & Ert, E. (2007). Evaluating the reliance on past choices in adaptive learning models. *Journal of Mathematical Psychology, 51*, 75–84.

Yechiam, E., Stout, J. C., Busemeyer, J. R., Rock, S. L., & Finn, P. R. (2005). Individual differences in the response to foregone payoffs: An examination of high functioning drug abusers. *Journal of Behavioral Decision Making, 18*, 97–110.

## Appendix A: A detailed procedure for the one-step-ahead prediction method

1. For a given agent (participant) $i$, a model is provided with all parameter values (e.g., $A$, $W$, and $c$).
2. In the beginning, it is assumed that $E_j(t = 0) = 0$ for any deck $j$. Therefore, $Pr[D(1) = j] = 0.25$ for each deck $j$ (Equation 5).
3. On trial $t = 1$, the model is provided with $Y_i(t = 1)$, the deck the agent actually chose and $X_i(t = 1)$, actual payoffs experienced on the trial. Then, compute $u(t)$ (Equations 1 or 2).

4. The expectancies are updated by Equation 3 or 4.
5. The probability that deck $j$ will be chosen on the next trial ($Pr[D(t+1)=j]$) is updated accordingly by Equation 5.
6. Repeat Steps 3 through 5 up to and including $t = n-1$ where $n$ is the total number of trials (= 120 in this study).
7. Compute log likelihood (LL):

$$\mathrm{LL}_{\mathrm{Model,i}} = \sum_{t=1}^{n-1}\sum_{j=1}^{4} \ln(Pr[D_j(t+1)=j)\cdot\delta_j(t+1)]),$$

   where $\delta_j(t+1)$ is a dummy variable, which is 1 if deck $j$ is chosen on trial $t+1$ and 0 otherwise.
8. Compute $\mathrm{LL}_{\mathrm{baseline}}$. Note that the baseline model is a random model.
9. Compute $G^2$ (Equation 9).

## Appendix B: The recipe for the simulation method

This is to illustrate the steps of the simulation method so that readers can more easily grasp the difference between the one-step-ahead prediction method and the simulation method. Be reminded that the procedure for the simulation method is quite similar to that of the one-step-ahead prediction method, but they are different in that the simulation method does not use observed choices of participants and makes predictions for the entire sequence of all choices.

To simulate data of an individual $i$ generated by a model, the following steps are taken:

1. For a given agent (participant) $i$, the model is provided with all parameter values (e.g., $A$, $W$, and $c$).
2. In the beginning, it is assumed that $E_j(t) = 0$. Therefore, $Pr[D(1)=j] = 0.25$ for each deck $j$ (Equation 5).
3. On trial $t = 1$, randomly generate a number from a uniform distribution that is between 0 and 1.
4. Using the number, a choice is generated on trial $t = 1$. The probability of choosing a deck $j$ on trial $t = 1$ is proportional to its $Pr[D(1)=j]$ (i.e., on Trial 1, all decks have equal probabilities [= 0.25] of being chosen because $Pr[D(1)=j] = 0.25$ for deck $j$). Note that, in contrast to the simulation method, the one-step-ahead prediction method "knows" the actual choice on the current trial of the agent and uses the information to do the following steps.
5. The choice yields some amount of gain or loss for a given task. The utility of the choice is calculated by Equation 1 or 2.
6. The expectancies are updated by Equations 3 or 4.
7. The probability that deck $j$ will be chosen on the next trial ($Pr[D(t+1)=j]$) is updated accordingly by Equation 5.
8. Repeat Steps 3 through 7 up to and including $t = n-1$ where $n$ is the total number of trials (= 120).

9. Repeat Steps 2 through 8 100 times for each agent and calculate $\overline{Pr_i}[D(t) = j]$, which is the averaged probability across 100 stimulations of the agent $i$.
10. Repeat Steps 1 through 9 for all agents ($N = 36$ in this study) and compute $\bar{D}_{sim,j}(t)$:

$$\bar{D}_{sim,j}(t) = \frac{1}{N}\sum_{i=1}^{N}\overline{Pr_i}[D(t) = j].$$

.
11. Compute mean square deviation scores (Equation 11).