Journal of Behavioral Decision Making, *J. Behav. Dec. Making* (2010)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/bdm.722

Instance-based Learning: A General Model of Repeated Binary Choice

TOMÁS LEJARRAGA*, VARUN DUTT and CLEOTILDE GONZALEZ

Dynamic Decision Making Laboratory Department of Social and Decision Sciences, Carnegie Mellon University, PA, USA

ABSTRACT

A common practice in cognitive modeling is to develop new models specific to each particular task. We question this approach and draw on an existing theory, instance-based learning theory (IBLT), to explain learning behavior in three different choice tasks. The same instance-based learning model generalizes accurately to choices in a repeated binary choice task, in a probability learning task, and in a repeated binary choice task within a changing environment. We assert that, although the three tasks are different, the source of learning is equivalent and therefore, the cognitive process elicited should be captured by one single model. This evidence supports previous findings that instance-based learning is a robust learning process that is triggered in a wide range of tasks from the simple repeated choice tasks to the most dynamic decision making tasks. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS instance-based learning; repeated choice; decisions from experience; model comparison; probability learning; model generalization

INTRODUCTION

A common approach in the study of decision making involves observing human performance in a choice task followed by the development of a cognitive model that reproduces that behavior and predicts new unobserved behavior within the same task (Cassimatis, Bello, & Langley, 2010). Usually, new tasks lead to the design of new models, resulting in highly task-specific models that fail to reproduce behavior even in closely related tasks. The efficacy of this approach, therefore, has been questioned (Anderson & Lebiere, 2004; Cassimatis et al., 2010; Newell, 1973). A particular field in which model specificity prevails is repeated binary choice: Different models have been proposed to explain behavior in tasks that only differ slightly. For example, different models are used to predict learning in repeated binary choice with outcome feedback, with feedback about foregone payoffs, and probability learning (one exception has been RELACS by Erev and Barron, 2005).

Like other researchers, we also question the value of models that are too task-specific. However, we assert that the cognitive process underlying learning is more often triggered by the source of information observed by the decision maker than solely by the nature of the task. For example, while probability learning and repeated binary choice are different tasks, the source for learning is equivalent: Decision makers rely on exploration of the alternatives and their resulting outcomes, which are drawn sequentially from binomial distributions. Therefore, the same model should be used to explain behavior in different tasks that yield similar information for learning.

In this paper, we present a model that predicts behavior in three different choice tasks that provide the same type of information for learning. The model relies on instance-based learning theory (IBLT), originally proposed to describe decision making in complex dynamic decision making tasks (Gonzalez, Lerch, & Lebiere, 2003).

Next, we describe IBLT and then an instance-based learning model (hereafter "IBL model") for binary choice. We then explain how the same IBL model makes relatively accurate predictions of learning from experience in three different tasks: a repeated choice task, a probability learning task, and a repeated choice task in non-stationary environments. We conclude by suggesting that IBLT is a robust representation of decisions from experience.

INSTANCE-BASED LEARNING THEORY

IBLT was developed to explain decision-making behavior in dynamic tasks (Gonzalez & Lebiere, 2005; Gonzalez et al., 2003). In dynamic tasks, individuals make repeated decisions attempting to maximize gains over the long run (Edwards, 1962; Rapoport, 1975). According to Edwards (1962), dynamic decision tasks are characterized by decision conditions that change spontaneously and with time, inaction, and as a result of previous decisions.

Edwards (1962) proposed an initial taxonomy of dynamic tasks, ranging from the least dynamic tasks, where actions are sequential in an environment that is constant and where neither the environment nor the individual's information about the environment is affected by previous decisions; to the most dynamic tasks, where the environment and the individual's information about it changes over time and as a function of previous decisions. This taxonomy was later extended to include an even more dynamic characteristic in Edwards' taxonomy: that decisions are made in real time, and thus their outcomes depend on the time at which the decision is made (Brehmer, 1992; Hogarth, 1981).

IBLT was proposed as a descriptive account of the cognitive structures and learning processes in dynamic decision making. IBLT characterizes learning in dynamic

* Correspondence to: Tomás Lejarraga, Dynamic Decision Making Laboratory, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
E-mail: tleja@andrew.cmu.edu

tasks by storing in memory a sequence of action-outcome links (called ''instances'') produced by experienced events. The theory proposes a generic decision-making process that involves: recognition, judgment, choice, execution, and feedback as the main steps of the process that allows decisions to improve as experience accumulates in memory and according to the variability of the environment. The following are some of the components that IBLT assumes (see Gonzalez et al., 2003, for a detailed description of the theory):

- *Instances*: Instances are examples of previous choices that are encoded in memory. Each instance contains cues about the situation in which a decision was made, the decision itself, and the subsequent outcome. Situational cues are relevant in dynamic environments because situations change and not all experiences are informative for future choice situations.
- *Activation*: Learning resides in a mechanism called activation (Anderson & Lebiere, 1998), which relies on the frequency and recency of experienced choices and outcomes. IBLT assumes that the instances experienced by the decision maker are activated in memory as a function of the aspects of their previous occurrence: More recent and frequent instances are more active in memory than less recent and less frequent ones.
- *Similarity*: Choice situations are never equivalent in dynamic tasks as environments change over time. Thus, past experiences are not necessarily informative in new conditions. A similarity rule, defined on situational cues, evaluates the resemblance of previous situations to the current situation, so that the model only considers experiences that occurred in similar situations.
- *Blending*: IBLT uses a mechanism originally proposed by Lebiere (1999) to assess the attractiveness of the alternative options based on the observed outcomes in previous similar situations. For each option, the values of all observed outcomes resulting from selecting that option are blended into a single blended value. The nature of blending is similar to that of combining utilities (or values) and probabilities (or decision weights) in expected utility theory (von Neumann & Morgenstern, 1947) or prospect theory (Kahneman & Tversky, 1979).

IBLT was originally implemented within the ACT-R cognitive architecture (Anderson & Lebiere, 1998) to account for decision making in a complex dynamic task within a highly dynamic environment (Gonzalez et al., 2003). ACT-R is both a theory of human cognition and a programming environment. As a programming environment, it allows researchers to design tasks and create models that incorporate ACT-R's theory of cognition. After IBLT was proposed, other models based upon the theory have been developed within the ACT-R architecture and demonstrated a close approximation to human decision making in multiple tasks (e.g., Gonzalez & Lebiere, 2005; Lebiere, Gonzalez, & Martin, 2007; Martin, Gonzalez, & Lebiere, 2004). Moreover, models based on IBLT and within ACT-R have been implemented to account for repeated choices (Lebiere et al., 2007; Stewart, West, & Lebiere, 2009). The diversity of models based on IBLT that closely represent human behavior suggests that IBLT is a generic theory of decisions from experience (Gonzalez & Dutt, 2010). In fact, one indication of the potential of IBLT as a general theory of decisions from experience is the observation that the winner of a recent choice prediction competition (the Technion Prediction Tournament; Erev et al., 2010) is an ACT-R model based on IBLT. However, these implementations are often task-specific and rarely demonstrate that the same model can account for behavior in multiple tasks.

In part, the generalization of IBLT may have been limited by the complexity of the ACT-R architecture. The ACT-R architecture has evolved to be highly complex, often allowing modelers the freedom to adopt new approaches in the representation of the learning process. Additionally, the development of IBL models within the ACT-R architecture requires considerable technical knowledge. Such an approach, although convenient for complex tasks, is not parsimonious for simple tasks like repeated choices, and hinders model generalization and reuse.

In the current research, we present a simple model based on IBLT, and show that it accounts for choice behavior in simple dynamic tasks (i.e., according to the taxonomy proposed by Edwards, 1962). Given the simplicity of repeated choice tasks, we developed a simple version of an IBL model (in Microsoft Excel) to demonstrate the generality of the learning process in multiple tasks. In the following section, we describe this model for repeated binary choice based on the mechanisms proposed in IBLT.

AN INSTANCE-BASED LEARNING MODEL FOR REPEATED BINARY CHOICE

In repeated binary choice, the model chooses one of two options by selecting the one with the highest *blended value V* (Lebiere, 1999). The blended value of option j (e.g., a gamble that pays \$4 with .8 probability or \$0) is defined as

$$V_j = \sum_{i=1}^{n} p_i x_i \qquad (1)$$

where x_i is the value of the observed outcome i (e.g., either \$4 or \$0, in the previous example) and p_i is the *probability of retrieval* of that outcome from memory. Because x_i is the value of the observed outcome, the number of terms in the summation changes when new outcomes are observed within option j. Thus, $n = 1$ if j is a safe option with one possible outcome. When j is a risky option with n possible outcomes, $n = 1$ when one of the outcomes has been observed, $n = 2$ when the two different outcomes have been observed, and so on. Given that the choice rule entails selecting the option with the highest blended value, the model is naturally suited for choices among more than two options.

As Equation (1) shows, the blended value of an option is the sum of all the observed outcomes x_i weighted by their probability of retrieval. At any trial t, the probability of retrieval of observed outcome i is a function of the activation of that outcome relative to the activation of all the observed

DOI: 10.1002/bdm

outcomes within that option, given by

$$P_{i,t} = \frac{e^{A_{i,t/\tau}}}{\sum_j e^{A_{j,t/\tau}}} \tag{2}$$

where τ is random noise defined as $\tau = \sigma \cdot \sqrt{2}$, and σ is a free noise parameter. Noise in Equation (2) captures the imprecision of recalling past experiences from memory. The activation of each outcome in memory depends upon a mechanism from ACT-R (Anderson & Lebiere, 1998). The activation of an outcome in a given trial is a function of the frequency of its occurrence and the time past since each of these outcomes occurred. At each trial t, *activation A* of outcome i is

$$A_{i,t} = \sigma \ln\left(\frac{1-\gamma_{i,t}}{\gamma_{i,t}}\right) + \ln \sum_{t_p \in \{1,\ldots,t-1\}} (t-t_p)^{-d} \tag{3}$$

where d is a free decay parameter; $\gamma_{i,t}$ is a random draw from a uniform distribution bounded between 0 and 1; and t_p is each of the previous trial indexes in which the outcome i was observed. The first term in the right-hand side of Equation (3) adds a random error to the activation process, and it is intended to represent the noise associated with memory activation. The summation will include a number of terms that coincides with the number of times that outcome i was observed in previous trials. Therefore, the activation of an outcome increases with the frequency of observation (i.e., by increasing the number of terms in the summation) and with the recency of those observations (i.e., by small differences in $t-t_p$). The decay parameter d affects the activation of the outcomes directly, as it captures the rate of forgetting. A low decay value translates into higher activations in memory, which implies longer lasting memory.

Because memory is unlikely to be empty when starting a task, the model assumes that some initial expectation exists in memory before any choice is made. One initial outcome for each option is pre-populated in memory and may represent the initial payoff expectations that participants bring to the laboratory. These values are set to be higher than those observed in the gamble in order to trigger exploration between options in early choices. Because these outcomes are never observed, they are never reinforced and therefore their activation decays in memory and soon becomes trivial. These initial outcomes are not model parameters.

When choices yield feedback about foregone outcomes in addition to obtained outcomes, the model activates the observed outcomes of the non-chosen option in the same manner as if they were the outcomes of a chosen option. Notably, the same activation (Equation (3)) process affects obtained and foregone outcomes. The model, thereby, is suited to process foregone payoffs naturally.

Table 1 shows how the IBL model works for a choice between options A and B. Option A is a risky prospect that offers $4 with a .8 probability or $0 otherwise, and option B is a safe prospect that offers $3 with certainty. The table shows a simulation for 10 trials, as indicated in the leftmost column. The second column shows the choice between A and B, derived from comparing the blended value of each option displayed in the two rightmost columns. The 3rd and 4th columns show the outcomes observed by the model in each trial. The following set of columns shows the activation of each of the observed outcomes and a pre-populated initial expectation set arbitrarily at $30. The following set of columns shows the probability of retrieval of each observed outcome and the blended values.

In trial 1, none of the possible outcomes ($4 and $0 for option A, or $3 for option B) are active in memory because the hypothetical subject is not aware that they exist. The only active outcome is $30, the initial expectation for each option. The initial expectations are not assumed to be known by the subjects, but they are arbitrarily pre-populated in memory to trigger exploration in initial trials. As a result, the probability of retrieval of each initial expectation is 1, given that it is the only outcome active in memory for each option. Thus, the model chooses randomly in the first trial. The observed outcome of choosing A in trial 1 is $4, as shown in column 3.

Table 1. Application of the IBL model on a sample choice problem between a risky option ($4 with .8 probability or $0 otherwise) and a safe option ($3 for sure)

				Activation (Equation (3))					Probability of retrieval (Equation (2))					Blended values (Equation (1))	
				Option A			Option B		Option A			Option B			
Trial	Choice	Outcomes for Option A	Outcomes for Option B	$4	$0	$30	$3	$30	$4	$0	$30	$3	$30	Option A	Option B
1	A	4				−0.90		−0.78	0.00	0.00	1.00	0.00	1.00	30.00	30.00
2	B		3	−4.46		−7.90		−1.67	0.84	0.00	0.16	0.00	1.00	8.29	30.00
3	A	0		−4.83		−6.99	1.98	−4.97	0.73	0.00	0.27	0.96	0.04	10.91	3.99
4	B		3	−5.13	−0.89	−9.37	−3.09	−7.81	0.12	0.87	0.02	0.90	0.10	0.94	5.64
5	A	4		−5.06	−5.09	−4.88	−2.48	−4.02	0.33	0.32	0.35	0.67	0.33	11.92	11.80
6	B		3	0.11	−5.31	−6.95	−4.02	−4.70	0.90	0.07	0.03	0.58	0.42	4.56	14.36
7	A	4		−3.28	−8.46	−8.78	−3.02	−8.74	0.86	0.07	0.06	0.94	0.06	5.38	4.71
8	A	4		−2.08	−8.64	−8.70	−1.83	−10.77	0.92	0.04	0.04	0.99	0.01	4.88	3.39
9	A	4		−3.45	−10.60	−13.16	−6.46	−14.00	0.96	0.03	0.01	0.97	0.03	4.12	3.75
10	A	4		−0.56	−9.21	−17.19	−10.29	−17.82	0.98	0.02	0.00	0.97	0.03	3.94	3.75

Note: The parameters used in this example are $d = 5$ and $\sigma = 1.5$. These values were obtained by fitting the model to experimental data from 60 problems in the TPT's estimation set.

DOI: 10.1002/bdm

In trial 2, then, \$4 is active in memory (activation $= -4.46$). Notice that the probability of retrieval for the outcomes in option A now becomes distributed between \$30 and \$4 according to Equation (2): \$4 receiving a larger portion of the weight than \$30 because it has been observed more recently. Further notice that the activation of \$30 decays in both options, and it does so through the end of the task. The observation of \$4 in trial 1 altered the balance in the blended values, which now favors option B over A (i.e., the \$30 expectation is promoting exploration in these initial trials). Choosing B yields an observed outcome of \$3, which now becomes active in memory and distributes the probability of retrieval of outcomes within option B. The process continues successively.

Next, we demonstrate the performance of this IBL model in three different tasks: repeated binary choice in the Technion Prediction Tournament (TPT), probability learning, and repeated choice with non-stationary probabilities.

THE TECHNION PREDICTION TOURNAMENT

The TPT (Erev et al., 2010) was a competition in which different models were submitted to predict choices made by experimental participants. Competing models were evaluated following the generalization criterion method (Busemeyer & Wang, 2000), by which models were fitted to choices made by participants in 60 problems (the estimation set) and later tested in a new set of choices in 60 problems (the test set) with the parameters obtained in the estimation set. The 120 problems involved the choice between a safe option that offered a medium (M) payoff with certainty and a risky option that offered a high (H) payoff with some probability (pH) and a low (L) payoff with the complementary probability, just like the example used in Table 1. M, H, pH, and L were generated randomly, and a selection algorithm assured that the 60 problems in each set differed in domain (positive, negative, and mixed payoffs) and probability (high, medium, and low pH). The resulting set of problems was large and representative. The characteristics of the 60 problems in the test set are included in Figure 2.

For each of the 60 problems, a sample of participants (100 for the problems in the estimation set and 160 for the problems in the test set) was randomly assigned into five groups, and each group completed 12 of the 60 problems. Each participant was instructed to select between two unlabeled buttons on a computer screen for an unspecified number of trials. One button was associated with a risky option and the other button with a safe option. Each choice yielded an outcome in Sheqels, and the outcome of the non-chosen option was not presented. For a more detailed description of the experimental procedure, please see Erev et al. (2010, p. 21).

IBL model performance in the Technion Prediction Tournament

The IBL model described above was evaluated under the same conditions as the other models in the TPT. Following the generalization criterion method used in the competition, we fitted the IBL model to the 60 problems in the estimation set. We calibrated two free parameters, noise σ and decay d, to minimize the mean squared distance (MSD) between the observed proportion of risky choices (R-rate) across problems and the R-rate made by 100 simulated participants generated by the IBL model. MSD was minimized at .0056 when $\sigma = 1.5$ and $d = 5$. The fitted model yielded a correlation of $r(58) = .9060$ ($p < .01$) between participants' and the model's R-rate. Then, we evaluated the IBL model against choices made on the 60 problems of the test set with the parameters obtained in the estimation set.

The IBL model predicted choices accurately. Figure 1 shows that in the estimation and test sets, the IBL model outperformed all competing models in the E-repeated condition of the TPT, including an instance-based learning model called ACT-R with sequential dependencies and blending memory (submitted by Stewart, West, and Lebiere and reported in Erev et al., 2010, hereafter ''ACT-R model'') that won the competition. The results show that IBL's MSD is the lowest and the correlation is the highest among competing models.

It is important to keep in mind that IBL has two free parameters, the same number of parameters as the winning ACT-R model, which compares favorably to the four parameters in the explorative sampler with recency (ESR, Erev, Ert, & Yechiam, 2008; Erev et al., 2010), the model used as a benchmark and that provided the best predictions in the TPT. Figure 2 shows the learning curves of all the 60 problems in the test set of the competition. As can be seen, the IBL model accurately predicted learning in most of the problems.

The performance of the IBL model in the 60 problems of the test set was analyzed as a function of the characteristics of the problems. The problems were separated by domain into positive, negative, and mixed, and by the probability of the high outcome into high, medium, and low. The model predictions yield no marked pattern of error, so the small prediction error can be attributed to the inherent randomness of the experimental data.

The current analysis extends the findings in the TPT in two ways. First, it examines choice behavior over trials, whereas the original article focused on the prediction of the aggregate choice rate across trials (Erev et al., 2010). Second, it proposes a fundamental simplification of the winning ACT-R model, discussed next.

We obtained the ACT-R model from Stewart et al. (2009), and we generated predictions from our IBL model and the ACT-R model for comparison. Although the IBL model and the ACT-R model are based on IBLT, they differ in many important psychological assumptions. First, the ACT-R model assumes strong sequential dependencies: Instances stored in memory include a choice, its outcome, and the previous two choices that preceded the current choice. To describe this process, let us refer to the previous example (\$4 with a .8 probability or \$0 otherwise and \$3 with certainty). Given a hypothetical sequence of choices and observed outcomes R $\rightarrow$ \$4, S $\rightarrow$ \$3, R $\rightarrow$ \$4, S $\rightarrow$ \$3, R $\rightarrow$ \$4 (R for risky and S for safe) in the first five trials, the ACT-R model

DOI: 10.1002/bdm

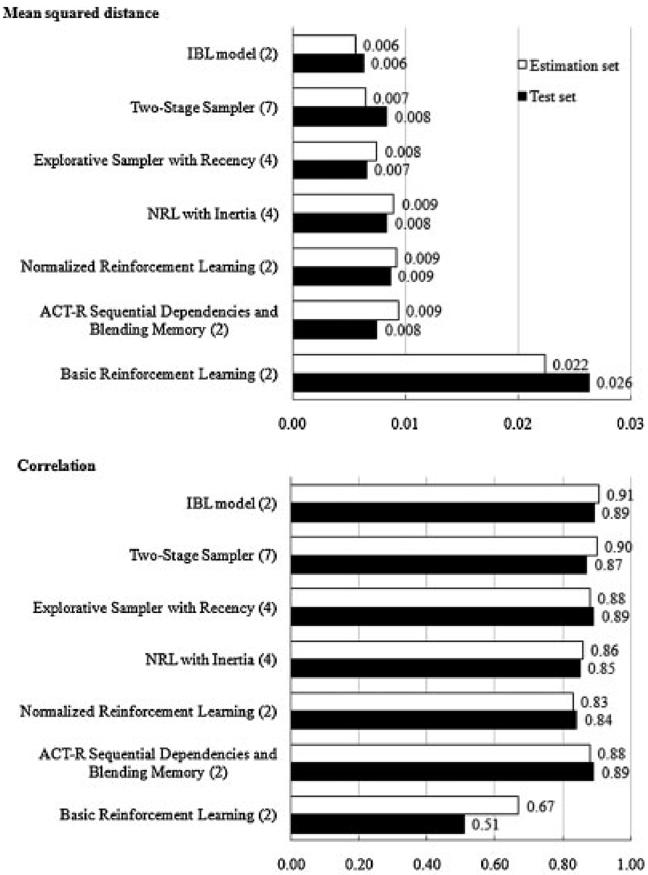

Figure 1. Scores of IBL and the models that participated in the Technion Prediction Tournament. The upper graph shows the mean squared distance between observed R-rate and each model's prediction. The lower graph shows the correlation between the observed R-rate and each model's prediction. The number of free parameters of each model appears in parenthesis after each model name

would begin to store instances from trial 3 onwards. In trial 3, the model would store an instance that records the risky choice, the \$4 obtained and the two choices leading to that choice: S (in trial 2) and R (in trial 1). In trial 4, a new instance would be created that stores the safe choice, the \$3 obtained, and the two preceding choices R and S. In trial 5, upon observing the \$4 outcome, the model would store a new instance even though a \$4 outcome had already been observed in trial 3. The reason that the model stores these two \$4 outcomes as different instances is that they present different histories (i.e., different two preceding choices), and therefore, the model assumes they are different experiences.

Second, the ACT-R model considers only stored instances whose history matches the current history for each choice. By recalling the instances with matching histories, the ACT-R model assumes the following logic: If a given choice sequence has led to high outcomes in the past, it will also lead to high outcomes in the current choice which matches that history. For example, in trial 6 of the previous example, the model would consider only instances whose history is a sequence of S and R in the exact order of occurrence. Then, each instance (with unique histories) is active in memory according to an activation equation that operates in the same way as Equation (3) used in the IBL model.

Third, the ACT-R model assumes that only the instances whose activation surpasses an arbitrary retrieval threshold will be successfully retrieved from memory.

The IBL model presented here, in contrast, does not store the history of choices nor does it assume a retrieval threshold. In our IBL model, every experience has a probability of being retrieved (P_i) from memory: Old outcomes have a lesser weight on the option's blended value (Equation (1)) than recent outcomes, but all experiences receive some weight. This distinction results in different blending mechanisms used between the two models. The ACT-R model assumes that the blended value of an option is the average of the values of observed outcomes (with unique histories) weighted by their activations, for those instances whose activation surpasses the retrieval threshold. In contrast, the IBL model assumes no retrieval threshold, and the blended value of an option is the average of the values of observed outcomes weighted by their probability of retrieval, as described in Equation (1).

The relatively accurate predictions of the ACT-R model, in addition to how the model treats the two latest choices, imply strong sequential dependencies in the observed choice behavior. The IBL model captures this behavior through a high decay parameter ($d = 5$ in IBL compared to .5 in the ACT-R model), which implies that outcomes are quickly forgotten and only the most recent outcomes receive a critical weight in decisions.

The 60 problems in the test set of the TPT were ranked according to the accuracy of the IBL predictions. Overall, the IBL model performs better than the ACT-R model in 34 of the 60 problems. Figure 3 shows the observed and predicted choices for nine problems that represent the deciles from the 1st to the 9th. In each of the nine graphs, we plotted the predictions by the IBL model and those by the winning model of the competition across the 100 trials.

Figure 3 shows the relative accuracy of both models across the 60 problems: Even in the problems with less accurate predictions (e.g., 1st through 3rd deciles), both models performed relatively well.

PROBABILITY LEARNING

Probability learning refers to the study of how individuals predict the outcome of two mutually exclusive, random events. In a typical probability-learning task, participants predict which of two lights will turn on in a number of trials. In the standard version of the task, the probability that a light turns on is unknown to participants, who learn it from experience. Early studies (Edwards, 1961) suggest a tendency of participants to choose the more likely event with a probability that is similar to the event probability, a phenomenon referred to as "probability matching." The source of information for learning in this task is the same as in the TPT task described above: an observed frequency of random binary outcomes. Because the source of information for learning is the same between the probability learning and

DOI: 10.1002/bdm

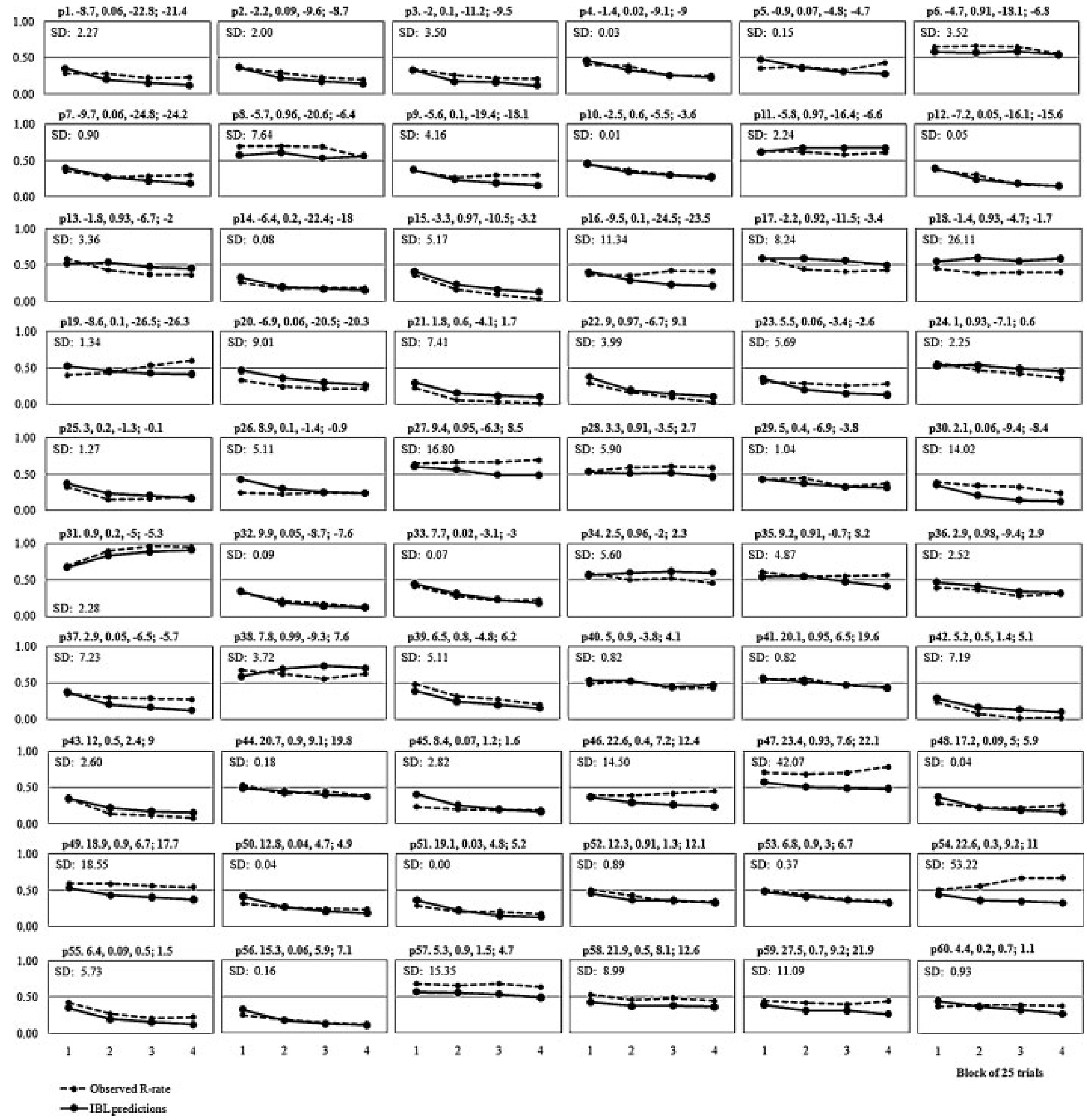

Figure 2. Mean observed proportion of risky choices (R-rate) and IBL predictions for each block of 25 trials in each of the 60 problems in TPT's test set. The title of each individual graph indicates the problem number followed by H, p(H), L, and M. For example, problem 1 involves a risky option that offers −8.7 with a 0.06 probability or −22.8 otherwise and a sure option of −21.4 with certainty. *SD* in each graph denotes the squared distance between observed R-rate and IBL predictions across the 100 trials

the repeated choice tasks, we hypothesize that the same IBL model presented above will predict probability learning accurately.

Myers, Reilly, and Taub (1961) examined probability-learning behavior in a set of 27 diverse problems (displayed in Table 2). They designed a 3×3 between-subjects experiment where problems varied in the frequency of occurrence of the two lights, the reward for correct predictions, and the cost for incorrect predictions. As a result, the problems covered a wide space of probability learning problems.

Participants in Myers et al. (1961) were assigned randomly to play one of the 27 problems. They had to predict, in each of 150 trials, which of two lights would turn on. Each participant was awarded 100 chips (worth .5¢ each) as game currency, and they could win additional chips by predicting correctly or lose chips by predicting incorrectly. The amount of chips earned at the end of the experiment was exchanged for money.

The frequencies of occurrence of the two lights were 90-10 (i.e., one light turned on 90% of the times and the other light turned on 10% of the times), 70-30, and 50-50. The amount of chips gained with each correct prediction depended on the light being correctly predicted. Because high frequency lights are easier to predict, correct predictions of high frequency lights were rewarded with fewer chips than correct predictions of low frequency lights. There were three gain ratios that determined the rewards: 1:4, 1:2, and 1:1. For example, in the 1:4 condition, correct predictions of low frequency lights were rewarded with 4 chips, while correct predictions of high frequency lights were rewarded with 1 chip. In the 1:1 condition, correct predictions were rewarded

DOI: 10.1002/bdm

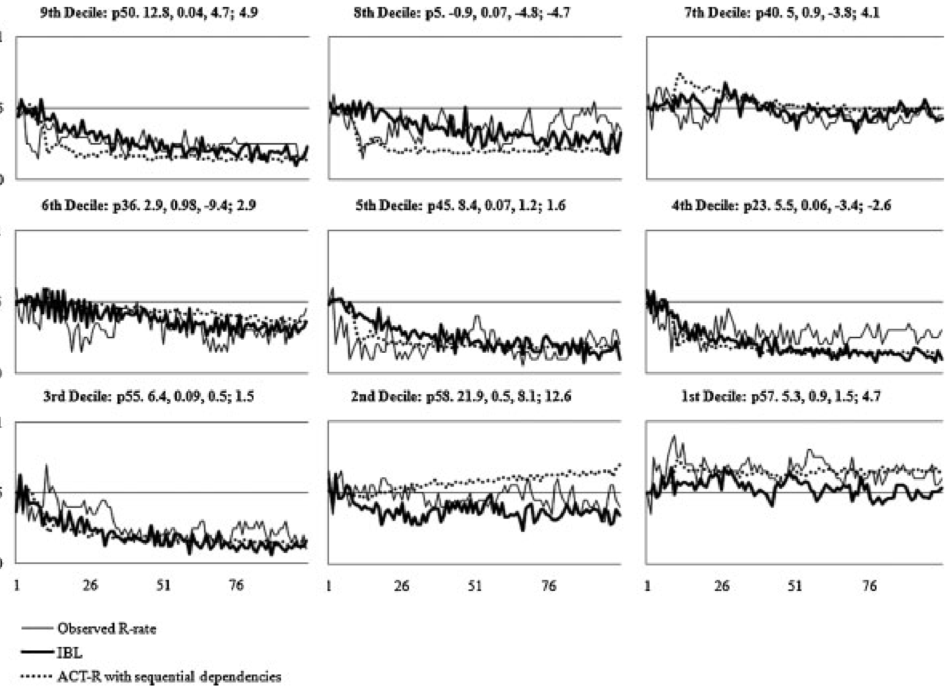

Figure 3. Mean observed R-rate and predictions by 100 trials by the IBL model and ACT-R with sequential dependencies. The title of each individual graph indicates the problem rank according to the accuracy of IBL predictions, followed by the problem number, and then followed by H, p(H), L, and M. For example, the first problem in the upper left ranks in the 9th decile (90% of the 60 problems lie below this problem in the IBL accuracy rank), is problem 50, and involves a risky option that offers 12.8 with a 0.04 probability or 4.7 otherwise and a sure option of 4.9 with certainty

Table 2. Probability learning problems in Myers et al. (1961) and prediction by RELACS and IBL

	A		B			A-rate		
Problem	High	Low	High	Low	*p*(high)	Observed	RELACS	IBL
1	1	−1	1	−1	.5	.55	.50	.49
2	1	−1	1	−1	.7	.80	.83	.72
3	1	−1	1	−1	.9	.96	.96	.84
4	1	−1	2	−1	.5	.35	.36	.35
5	1	−1	2	−1	.7	.63	.74	.63
6	1	−1	2	−1	.9	.96	.95	.83
7	1	−1	4	−1	.5	.33	.25	.24
8	1	−1	4	−1	.7	.46	.58	.48
9	1	−1	4	−1	.9	.86	.92	.79
10	1	−2	1	−1	.5	.28	.36	.40
11	1	−2	1	−1	.7	.76	.74	.64
12	1	−2	1	−1	.9	.91	.95	.82
13	1	−2	2	−1	.5	.23	.29	.27
14	1	−2	2	−1	.7	.60	.66	.57
15	1	−2	2	−1	.9	.90	.94	.81
16	1	−2	4	−1	.5	.34	.23	.20
17	1	−2	4	−1	.7	.54	.52	.44
18	1	−2	4	−1	.9	.92	.91	.78
19	1	−4	1	−1	.5	.11	.26	.25
20	1	−4	1	−1	.7	.65	.58	.50
21	1	−4	1	−1	.9	.91	.92	.79
22	1	−4	2	−1	.5	.18	.23	.19
23	1	−4	2	−1	.7	.47	.51	.42
24	1	−4	2	−1	.9	.89	.91	.77
25	1	−4	4	−1	.5	.16	.21	.14
26	1	−4	4	−1	.7	.33	.42	.33
27	1	−4	4	−1	.9	.82	.87	.73

Note: The prediction problems in Myers et al. (1961) are expressed as choices between gambles in the following form: In both options, the gambles offer a high amount with probability *p*(high) and a low amount otherwise. Both gambles depend on the same random draw that determines the realization of high.

DOI: 10.1002/bdm

with 1 chip irrespective of the frequency of the lights. Likewise, because high frequency lights are easier to predict, incorrect predictions in high frequency lights cost more than incorrect predictions in low frequency lights. The cost ratios for incorrect predictions followed the same ratios as the gains. In the 1:4 condition, incorrect predictions of high frequency lights cost 4 chips, while incorrect predictions of low frequency lights cost 1 chip. In the 1:1 condition, incorrect predictions cost 1 chip for both lights. When the two lights occurred with the same frequency (in the 50-50 condition), the light assigned a higher gain was also assigned a lower cost.

IBL model performance in the probability learning problems

Myers et al. (1961) reported the mean choices for one of the options (the A-option) across participants and between trials 101 and 150. The observed outcomes (x_i) for the IBL model were the number of chips either earned or lost with each prediction. For example, in problem 1 (see Table 2), the observed outcomes of the A-option were 1 and -1 chips, and those of the B-option were 1 and -1 chips. The same representation of these probability-learning problems was followed by Erev and Barron (2005). IBL predictions were compared to the same dependent measure reported by Myers et al. (1961).

The results displayed in Table 2 show an accurate fit of the IBL model to the 27 problems. Of particular importance is the fact that the predictions for the probability learning problems were obtained with the same model used in the TPT task, and without altering the free parameters estimated in the previous task. The MSD between the observed proportion of A-choices (A-rate) and that predicted by the IBL model is .009, and the correlation across problems is $r(25) = .973$, $p < .01$. Erev and Barron (2005) tested RELACS, a four-parameter model with pre-defined strategies, in the same 27 problems reported here. The MSD between the observed A-rate and that predicted by RELACS is .004, and the correlation is $r(25) = .976$, $p < .01$. Although RELACS performs slightly better than our IBL model, IBL proposes a more parsimonious account of probability learning.

RELACS assumes that decisions follow one of three strategies: A strategy that selects the option with the highest recent payoff (fast best reply); a strategy that moves from a random selection in early trials to a two-stage case-based reasoning; and a strategy that involves continuous but diminishing exploration. In RELACS, these strategies are used stochastically, and the probability that each strategy is used is updated according to the outcomes that they determine. Despite different assumptions and levels of complexity, IBL and RELACS produce similar predictions. It is unclear which of RELACS' strategies was responsible for the choice pattern in the 27 probability learning problems and therefore it is difficult to establish a more meaningful comparison of models. Yet, given the high recency in the experimental data suggested by IBL's accurate fit and its high decay parameter, it is likely that RELACS relied mainly on the fast best reply strategy.

The analysis of the predictions by problem suggests that the IBL model produces more conservative predictions when compared to the observed proportions and when compared to the predictions by RELACS. In general, the IBL model predicts choice proportions that are closer to 0.5 than those revealed by participants. More extreme choice proportions are expected in the probability learning problems than in the TPT problems because the former involve mutually exclusive events. Since only one of the two lights turns on in each trial, the participant observes the outcome of the predicted option and also the outcome of the non-chosen option. Observing foregone outcomes reduces the motivation to explore and therefore produces more extreme choice proportions. Although the IBL model captures foregone payoffs naturally, the high decay parameter ($d = 5$) suggests heavy reliance on recent outcomes, which implies more unstable learning and high alternation between options. Moreover, greater noise ($\sigma = 1.5$) associated with the activation process implies an additional propensity to explore. Overall, high decay and noise translate into less extreme predictions. It is reasonable to assume, then, that the availability of foregone payoffs eases recall of each outcome, because outcomes are observed more frequently, and so lower decay and noise parameters may have produced even more accurate predictions.

Another reason that may contribute to the conservative predictions produced by the IBL model is the observation by Myers et al. (1961) that a few subjects were utilizing a maximizing strategy at the end of 150 trials. Maximizers generate more extreme choice proportions, and are unlikely to follow the assumptions of the IBL model. Therefore, it is reasonable to expect that the IBL model is capturing the non-maximizing majority in Myers et al. (1961).

NON-STATIONARY PROBABILITIES

Rakow and Miler (2009) explored repeated choice in situations where outcome probabilities for one of two options changed over trials. In their Experiment 1, 40 participants made 100 repeated choices between two risky options in four problems. In all the problems, each of two options involved a positive and a negative outcome, so participants could win or lose money with each decision. The novelty in the problems studied by Rakow and Miler (2009) is that, for one of the options, the probability of the positive outcome remained constant across trials (i.e., the stationary option S), while this probability changed across trials in the other option (i.e., non-stationary option NS). Changes in the probabilities for the NS option were gradual: The probability changed .01 per trial and over 40 trials. For example, problem 1 involved a choice between S that offered 10 with a .7 probability or -20 otherwise, and NS that initially offered 10 with a .9 probability or -20 otherwise. From trials 21 to 60, the probability of 10 in NS reduced by .01 in each trial, such that the probability of 10 in trial 60 and onwards was .5. In all four problems, the change in the probability was of .01

DOI: 10.1002/bdm

per trial and after the 40 changing trials, the probability remained unchanged at .5. The characteristics of each of the four problems are displayed in the titles of the individual graphs in Figure 4. After each choice, participants observed the outcome of the chosen option as well as the outcome of the option not chosen (i.e., the foregone payoff).

The apparatus and procedures are carefully described in Rakow and Miler (2009). Their results show that participants adapted slowly to probability changes, a behavior that was not captured particularly well by the associative choice model fitted in that study (Bush & Mosteller, 1955).

IBL model performance in repeated choices with non-stationary probabilities

We obtained the experimental data from Rakow and Miler (2009) for the four problems of Experiment 1, and we generated predictions from our IBL model by generating 100 simulated participants. For use as a benchmark, we also generated predictions from the explorative sampler model with recency (ESR, Erev et al., 2008, 2010), which provided the best predictions in the E-repeated condition of the TPT (ESR was a baseline model provided to contestants of the TPT to compare with their competing models). The same IBL and ESR models with the parameters estimated in the TPT's estimation set were tested against Rakow and Miler's (2009) data set. The IBL model fits the observed data particularly well (Figure 4). The MSD between observed and predicted NS choices in the four problems were .0009 for the IBL model and .0175 for the ESR model. When the MSD is measured across trials and problems, IBL yielded a .013 and a correlation of $r(398) = .816$ ($p < .01$), while the ESR model yielded an MSD of .033 and a correlation of $r(398) = .497$ ($p < .01$).

The ESR as implemented in Erev et al. (2010) included a simplification that assumes all of the previous experiences are equally similar. Biele et al., (2009), however, implemented a variant of the ESR model (the Contingent Sampler Model, CSM) better suited for dynamic problems that assumes contingent sampling. We adapted CSM to account for foregone payoffs and used it as a benchmark in Rakow and Miler's (2009) problems. However, the fit of CSM to the observed data was poorer than ESR's (MSD = .041 and correlation $r[398] = .401$, $p < .01$) and therefore we do not report on CSM further. One reason why CSM underperformed in Rakow and Miler's (2010) problems is that CSM's contingent sampling is relevant in problems with a Markov structure, as those in Biele et al., (2009). In Rakow and Miler (2009), problems have less of a structure, as previous outcomes have no influence over future outcomes, and therefore contingent sampling impairs fast adaptation.

To make the ESR better suited for non-stationary problems, we modified the model by assuming that sampling in ESR is contingent on the last observed outcome. While the original ESR model assesses the value of each option by drawing a random sample of past experiences from memory (where the last outcome is always sampled), our variant assumes that this sample of experiences is only drawn from the last outcome. Therefore, our implementation of the ESR

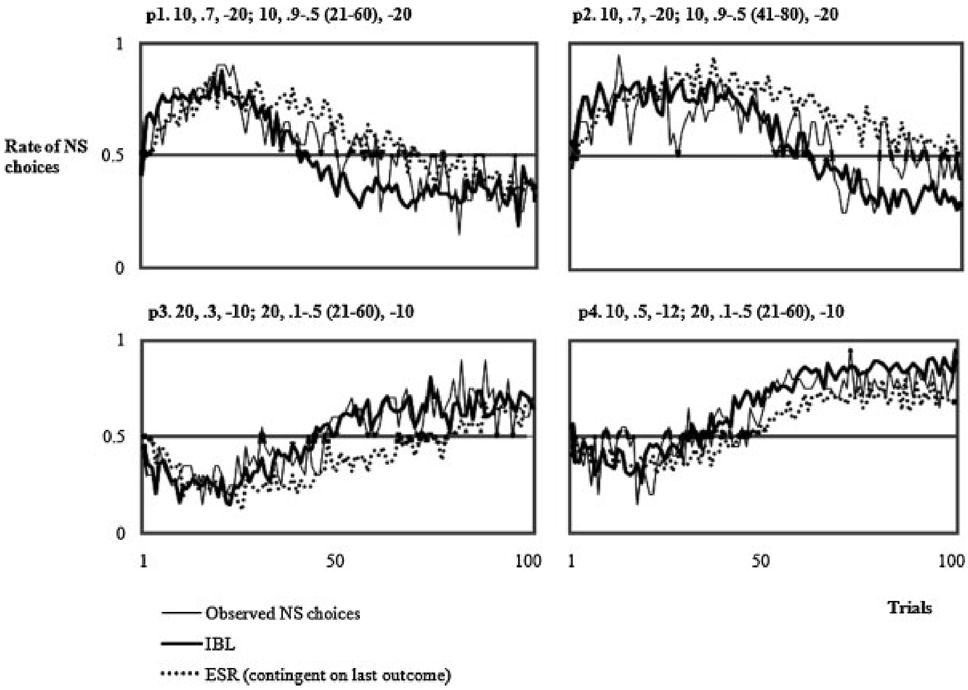

Figure 4. Mean observed choices for the non-stationary option (NS) and predictions by the IBL and ESR (with contingency on the last outcome) models across trials. The title of each individual graph shows the structure of the choice problem. For example, the upper left graph is problem 1, and involves a choice between a stationary option that offers 10 with a .7 probability or −20 otherwise, and a non-stationary option that offers 10 with a .9 probability or −20 otherwise. From trials 21 to 60, the probability of 10 in the non-stationary option is reduced by .1 in each trial, such that the probability of 10 in trial 60 and onwards is .5. The change in the probability is, for the four problems, of .1 per trial and for the 40 trials indicated between parentheses. After the 40 changing trials, the probability remains unchanged at .5, also for the four problems

DOI: 10.1002/bdm

implies a magnified recency effect. The new version of the ESR, with sampling contingent on the last outcome, fits the data better than the original ESR model, but still worse than IBL (See Figure 4). The MSD between observed and predicted NS choices in the four problems improved to .0077 and to .021 when MSD is measured across trials and problems. The new correlation between observed and predicted NS choices for the ESR contingent on the last outcome is $r(398) = .674$ ($p < .01$).

The relatively accurate predictions by the IBL model support the assertion that instance-based learning is an accurate representation of decisions from experience in choice tasks with non-stationary environments. Because the choice problems change gradually across trials, recent experiences are more informative than distant past experiences. In this environment, recency is an adaptive behavior. As Figure 4 shows, participants in Rakow and Miler's (2009) experiment adapted to changing conditions: Each of the observed learning curves shows a marked change in the trend of choices. This adaptation was captured by the IBL model quite well, and driven by the high decay parameter. Heavy reliance on recent outcomes is supported by the observation that the ESR model produces better predictions when contingent sampling is focused on the last observed outcome rather than on a subset of previous outcomes.

GENERAL DISCUSSION

Taken together, these studies show that IBLT presents an accurate and robust representation of decisions from experience in repeated binary choice. We assert that, as the three tasks provide similar types of information for learning, the cognitive process should not differ among them. We showed that good approximations to choice behavior were obtained by using the same model (with the same parameters) across three types of tasks, precluding the use of specific models in each task and promoting the reuse of existing models.

The results generalize the application domain of IBLT, and show that IBLT is a robust theory, well-suited for repeated choice tasks. This ability is illustrated by the precision of the model's predictions. Moreover, these results portray IBLT as a good candidate to reproduce human learning in other non-stationary environments; for example, in tasks where outcomes or probabilities vary as a function of previous choices, previous outcomes, or time. The study of decision making in such scenarios would help us understand the cognitive processes that allow individuals to adapt to changing environments, and would allow us to test existing models further. Given that IBLT was initially proposed for complex, real-time dynamic tasks, we can predict the roadmap needed to bridge the gap between simple choice tasks and more realistic tasks, by gradually increasing the dynamic characteristics of the repeated choice task as suggested by the taxonomy of dynamics proposed by Edwards (1962).

The model presented in this paper is a simplified version of the more complex IBLT used for dynamic decision making tasks. Therefore, the model excludes mechanisms that are irrelevant in a binary choice task but that can be critical in more complex dynamic tasks. For example, more complex tasks with changing situations across decisions would require that the IBL model includes similarity-based inference, which would allow the model to retrieve instances from memory on the basis of its similarity with newly encountered situations. Situational cues can be stored as part of instances and coupled with a similarity rule, would allow an IBL model to evaluate previous instances that share similar situational cues with the current choice situation: The model would base its current decision upon experiences that occurred in similar environmental conditions. These and other more sophisticated mechanisms make IBLT highly suitable for predicting behavior in more complex dynamic environments.

Everyday life presents a variety of decision conditions that differ in multiple ways, and we are often uncertain about the characteristics of the environment we encounter, whether we are in a simple relatively static situation or in a complex dynamic environment. A commuter who chooses between driving to work and taking the train learns from experience the most appropriate means to arrive to work on time. This choice scenario is simple and its conditions are not likely to change much from one day to the next. When the commuter arrives to work, however, she encounters a highly dynamic environment, where her decisions are made within a changing marketplace, often influenced by random shocks and even her own actions (for research on the dynamic aspect of managerial decision making, see Nelson and Winter, 1982; and Teece and Pisano, 1994). It is optimistic, therefore, to learn that instance-based learning, a relatively simple psychological mechanism, can cope with a wide range of situations in a robust manner, from the simple repeated tasks to the more dynamically complex tasks.

ACKNOWLEDGEMENTS

This research was supported in part by a grant from the National Science Foundation (Human and Social Dynamics: Decision, Risk, and Uncertainty, Award 0624228) to Cleotilde Gonzalez. The authors thank Tim Rakow and Katherine Miler for providing experimental data for analysis. They also thank Terrence C. Stewart, Robert West, and Christian Lebiere for sharing the code of the ACT-R model, and Tim Rakow, Ido Erev, Jolie Martin, and anonymous reviewers for their insightful comments.

REFERENCES

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Anderson, J. R., & Lebiere, C. L. (2004). The Newell test for a theory of cognition. *Behavioral and Brain Sciences*, *26*, 587–601. DOI: 10.1017/S0140525X0300013X

Biele, G., Erev, I., & Ert, E. (2009). Learning, risk attitude and hot stoves in restless bandit problems. *Journal of Mathematical Psychology*, *53*, 155–167. DOI: 10.1016/j.jmp.2008.05.006

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, *81*, 211–241. DOI: 10.1016/0001-6918(92)90019-A

Busemeyer, J. R., & Wang, Y. (2000). Model comparisons and model selections based on the generalization criterion methodology. *Journal of Mathematical Psychology*, *44*, 171–189. DOI: 10.1006/jmps.1999.1282

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

Cassimatis, N. L., Bello, P., & Langley, P. (2010). Ability, breadth, and parsimony in computational models of higher-order cognition. *Cognitive Science*, *32*, 1304–1322. DOI: 10.1080/03640210802455175

Edwards, W. (1961). Probability learning in 1000 trials. *Journal of Experimental Psychology*, *62*, 385–394. DOI: 10.1037/h0041970

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors*, *4*, 59–73.

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

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., et al. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, *23*, 15–47. DOI: 10.1002/bdm.683

Erev, I., Ert, E., & Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making*, *21*, 575–597. DOI: 10.1002/bdm.602

Gonzalez, C., & Dutt, V. (2010). Instance-based learning models of training. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*, 2319–2323.

Gonzalez, C., & Lebiere, C. (2005). Instance-based cognitive models of decision making. In D. Zizzo , & A. Courakis (Eds.), *Transfer of knowledge in economic decision-making* (pp. 148–165). New York: Palgrave Macmillan.

Gonzalez, C., Lerch, F. J., & Lebiere, C. (2003). Instance-based learning in real-time dynamic decision making. *Cognitive Science*, *27*, 591–635. DOI: 10.1016/S0364-0213(03)00031-4

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, *90*, 197–217. DOI: 10.1037/0033-2909.90.2.197

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

Lebiere, C. (1999). Blending. In *Proceedings of the Sixth ACT-R Workshop*, Fairfax, VA.

Lebiere, C., Gonzalez, C., & Martin, M. (2007). Instance-based decision making model of repeated binary choice. In R. L. Lewis , T. A. Polk , & J. E. Laird (Eds.), *Proceedings of the 8th international conference on cognitive modeling* (pp. 67–72). Ann Arbor, MI: Taylor & Francis/Psychology Press.

Martin, M. K., Gonzalez, C., & Lebiere, C. (2004). Learning to make decisions in dynamic environments: ACT-R plays the beer game. In M. C. Lovett , C. D. Schunn , C. Lebiere , & P. Munro (Eds.), *Proceedings of the sixth international conference on cognitive modeling* (Vol. 420, pp. 178–183). Pittsburgh, PA: Lawrence Erlbaum Associates Publishers.

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. DOI: 10.1037/h0047103

Nelson, R., & Winter, S. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). New York: Academic Press.

Rakow, T., & Miler, K. (2009). Doomed to repeat the successes of the past: History is best forgotten for repeated choices with nonstationary payoffs. *Memory and Cognition*, *37*, 985–1000. DOI: 10.3758/MC.37.7.985

Rapoport, A. (1975). Research paradigms for studying decision behavior. In D. Wendt , & C. Vlek (Eds.), *Utility, probability, and human decision making* (pp. 349–369). Dordrecht-Holland: D. Reidel Publishing Company.

Stewart, T. C., West, R., & Lebiere, C. (2009). Applying cognitive architectures to decision making: How cognitive theory and the equivalence measure triumphed in the Technion Prediction Tournament. In A. Howes , D. Peebles , & R. Cooper (Eds.), *9th international conference on cognitive modeling—ICCM2009* (pp. 561–566). Manchester, UK.

Teece, D. J., & Pisano, G. (1994). The dynamic capabilities of firms: An introduction. *Industrial and Corporate Change*, *3*, 537–556.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

Authors' biographies:

Tomás Lejarraga is a Postdoctoral Fellow at the Dynamic Decision Making Laboratory at Carnegie Mellon University. His current research interests focus on individual and group judgment and decision making, particularly on decisions from experience and learning.

Varun Dutt is a PhD candidate at the department of Engineering and Public Policy at Carnegie Mellon University. He is also the Knowledge Editor of the English daily, Financial Chronicle. His current research interests focus on risk perceptions in environmental decision making, dynamic decision making, and modeling of human behavior.

Cleotilde Gonzalez is Associate Research Professor and Director of the Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University. She is affiliated faculty at HCII, CCBI, and CNBC. She is in the Editorial board of the Human Factors Journal and Associate Editor of the JCEDM.

Authors' addresses:

Tomás Lejarraga, Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University, PA, USA.

Varun Dutt, Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University, PA, USA.

Cleotilde Gonzalez, Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University, PA, USA.