Psychological Review
2012, Vol. 119, No. 4, 888–892

0033-295X/12/$12.00 DOI: 10.1037/a0028004

COMMENT

Two Distinct Exploratory Behaviors in Decisions From Experience: Comment on Gonzalez and Dutt (2011)

Thomas T. Hills
University of Warwick

Ralph Hertwig
University of Basel

Gonzalez and Dutt (2011) recently reported that trends during sampling, prior to a consequential risky decision, reveal a gradual movement from exploration to exploitation. That is, even when search imposes no immediate costs, people adopt the same pattern manifest in costly search: early exploration followed by later exploitation. From this isomorphism in costless and costly search, the authors concluded that the same cognitive mechanisms underlay the control of sampling in 2 experimental paradigms employed to investigate decisions from experience: the sampling paradigm implementing costless search and the repeated-choice paradigm implementing costly search. We show that this is a misinterpretation of the human data resulting from drawing inferences about cognitive processes from data aggregated across individuals. Because of an inverse relationship between sample size and the propensity to explore, aggregating across individuals produces a pattern where exploration is gradually replaced by exploitation. On an individual level, however, there is no general reduction in exploration in the sampling paradigm. We list ensuing problems for the instance-based learning model Gonzalez and Dutt presented to explain the similarities between sampling and repeated decisions from experience.

Keywords: decisions from experience, risky choice, exploration, exploitation, search

Many choice ecologies allow us to examine two options before making a final costly decision, for example, when we leaf through two crime novels before deciding to purchase one, or when we go to a wine tasting before we buy a case. However, there are also myriad choice ecologies that do not permit us to sample without cost, as when we choose to travel down one roadway instead of another. Risky decisions based on experience often occur in these two types of ecologies. Either we are allowed to sample freely from the environment before the costly (i.e., consequential) choice or, alternatively, we must pay for our experience by making consequential decisions from the start. In the first type of ecology, *exploration* (e.g., obtaining information) and *exploitation* (obtaining reward) are separate processes, and the agent's only initial objective is to explore the environment in order to find out which of his or her actions is most instrumental in obtaining future rewards. The final costly decision (e.g., purchasing the book or the wine) often follows the costless exploration at a time of the agent's choosing, and thus the agent is not required to find a balance between the opportunity costs of both objectives. Exploration in this environment is thus like testing the water without footing the bill. In the second type of ecology, the sampled outcomes simultaneously provide consequential outcomes and information to the agent, and thus the agent faces the exploration–exploitation dilemma: "The agent has to *exploit* what it already knows in order to obtain reward, but it also has to *explore* in order to make better action selections in the future" (Sutton & Barto, 1998, p. 4). For instance, on vacation and sans a restaurant guide, we can continue to go to the same brasserie (because it was good enough on the first day), thus risking to miss out on even better ones; or we can explore some others, taking the risk of having culinary experiences worse than the ones that we safely could have had at the first brasserie.

In the laboratory, these two types of choice ecologies and associated risky decisions from experience have been abstracted into what Hertwig and Erev (2009) have called, respectively, the *sampling paradigm* and the *partial-feedback paradigm* (Gonzalez & Dutt, 2011, refer to the latter as the *repeated-choice paradigm*; henceforth, we adopt their language in order to avoid confusion). The latter imposes the exploration–exploitation dilemma on the agent; the former does not (see Gonzalez & Dutt, 2011, p. 525, for a detailed description of the paradigms). Because of this difference, researchers have commonly proposed disparate cognitive models to account for the choices obtained in each of the paradigms (see Erev et al., 2010; Gonzalez & Dutt, 2011, pp. 528–530). Challenging this theoretical divide, Gonzalez and Dutt (2011) recently proposed an important theoretical framework, the instance-based learning (IBL) model, with the goal of explaining choice and search in both paradigms using the same cognitive mechanisms (for a detailed description, see Gonzalez & Dutt, 2011, pp. 526–527).

Thomas T. Hills, Department of Psychology, University of Warwick, Coventry, England; Ralph Hertwig, Cognitive and Decision Sciences, University of Basel, Basel, Switzerland.

This research was supported by grants from the Swiss National Science Foundation to Thomas T. Hills (100014 130397/1) and Ralph Hertwig (100014 126558). We thank Laura Wiles for editing the article.

Correspondence concerning this article should be addressed to Thomas T. Hills, University of Warwick, Department of Psychology, Gibbet Hill Road, Coventry CV4 7AL, England. E-mail: t.t.hills@warwick.ac.uk

The IBL model builds in some important respects on the ACT–R (adaptive control of thought–rational) cognitive architecture (Anderson & Lebiere, 1998, 2003). Specifically, it assumes that a choice (given that the previous choice is not automatically repeated) represents the selection of the option with the highest utility (blended value). An option's blended value is a function of its associated outcomes and the probability of retrieving corresponding instances from memory. Memory retrieval depends on memory activation, which, in turn, is a function of the recency and frequency of the experience. The IBL model is particularly attractive because in the sampling paradigm it, unlike many other models that have been proposed, "predicts not only the final consequential choice but also the sequence of sampling selections" (Gonzalez & Dutt, 2011, p. 529), and because it offers a single learning mechanism (leading up to an instance's activation) that underlies both sequential choice and process behavior in the sampling and the repeated-choice paradigms.

The goal of the present comment is to show that Gonzalez and Dutt's (2011) commendable attempt to explain both paradigms in terms of the same mechanisms faces several potentially serious problems. We hope that their framework may ultimately be viable, but in its current form the IBL model appears to invite inaccurate inferences about individual search behavior, and the dynamic of exploration and exploitation across both experimental paradigms.

Is the Exploration–Exploitation Dynamic Indeed the Same in Both Paradigms?

The key issue concerns the relationship between exploration and exploitation in both paradigms. On the basis of their analyses, Gonzalez and Dutt (2011) concluded that what seemingly separates the paradigms might in fact unite them, the temporal dynamics of exploration and exploitation. Specifically, they interpreted their results to suggest that "in both paradigms, the A-rate decreases over an increased number of samples or trials. Furthermore, the same IBL model calibrated in one paradigm with the same parameters predicts the A-rate of the other paradigm" (pp. 538–539). The A-rate, or alternation rate, is a measure of the amount of exploration, and denotes the proportion of times that an individual moves from choosing one option to choosing the other option during periods of sampling (in the sampling paradigm; see also Hills & Hertwig, 2010) and repeated choices (in the repeated-choice paradigm). This measure of exploration is commonly used and is typically found to decay over time in repeated-choice tasks as individuals move from exploration to exploitation (e.g., Yechiam, Busemeyer, Stout, & Bechara, 2005). According to Gonzalez and Dutt's empirical and IBL model analyses (see their Figures 2 and 3), respondents in the sampling paradigm behave like respondents in the repeated-choice paradigm: Over time, exploitation supersedes exploration, and thus people at this nonconsequential search stage eventually act as if they had adopted an exploitative goal.

This isomorphism in nonconsequential search and consequential choice would indeed draw both paradigms nearer. However, the move from exploration to quasi-exploitation in the search phase of the sampling paradigm mischaracterizes what many individuals actually do. In a nutshell the problem is the following: Individuals in the sampling paradigm do not appear to show a decrease in A-rate over time. In the aggregate they do show this trend because, first, individuals vary in their search length in the sampling paradigm, and, second, individuals with higher A-rate tend to have shorter search lengths. When A-rate is analyzed as a function of the proportion of individual search length, the sampling paradigm reveals, on average, a strikingly *constant* level of exploration over the entire duration of the exploratory sampling phase. We explain this in more detail below.

Individual Versus Average Explorative Behavior in the Sampling Paradigm: The Human Data

Gonzalez and Dutt (2011) analyzed two data sets obtained in the sampling paradigm: the data collected by Hertwig, Barron, Weber, and Erev (2004) and the data collected in the sampling condition of the Technion Prediction Tournament (TPT; Erev et al., 2010). Figure 1 replots the average decreasing A-rate that they observed in two studies using the sampling paradigm alongside the decreasing A-rate observed in the repeated-choice paradigm from the TPT. Plotting the human data of these two paradigms alongside one another suggests a striking similarity, with both sampling and repeated-choice paradigms showing a gradual decrease in A-rate over time. However, inferring from the pattern in the sampling paradigms that individuals commonly move from exploration to exploitation is wrong. The reason lies in the inverse relationship between search length and A-rate.

As shown in Figure 2 (left and middle panels), individuals' distributions of total sample sizes and A-rate are significantly long-tailed (Shapiro–Wilks tests are $p < .001$ for all variables). Therefore, we employed a log transformation to compute the correlations (Shapiro–Wilks tests after the log transformation are $p > .1$ for both variables in both data sets). The correlation coefficients between the log of total sample size and log A-rate reveal a significant negative correlation for both data sets (Hertwig et al., 2004, data: $r = -.38$, $t(48) = -2.82$, $p < .01$; TPT data: $r = -.54$, $t(78) = -5.72$, $p < .001$). The negative correlations across individuals for both data sets are depicted in Figure 2 (right panels).

Given this negative relationship, aggregating the A-rate across individuals for different absolute numbers of samples (as done in Figure 1) means that those individuals whose total sample size is relatively small (and whose A-rate is relatively high) drop out of the analysis, leaving only those people behind who sample more and alternate less. Consequently, the (possibly) erroneous impression arises that exploration is superseded by exploration on an individual level. One way to deal with the inverse relationship of total sample size and A-rate is to normalize the A-rate across participants by dividing each person's sample sequence into 10 bins (each bin consisting of 10% of the search trials), and then computing the A-rate, separately for each of the bins. This normalization computes the A-rate for a specific proportion of the trials, which results in the number of individuals in each bin being approximately the same (e.g., individuals who only ever take two samples appear in only the first and last bin).

Figure 3A plots the resulting average A-rate per bin in both data sets and the repeated-choice data set. Now the conclusion is that, on average, the A-rate is quite constant across the sampling sequences, and thus qualitatively different from the declining exploration rate obtained in the repeated-choice paradigm (see Gonzalez & Dutt, 2011, Figures 2A and 3A). We found the same constant average A-rate (using normalized search sequences) when we investigated several other data sets from the sampling paradigm

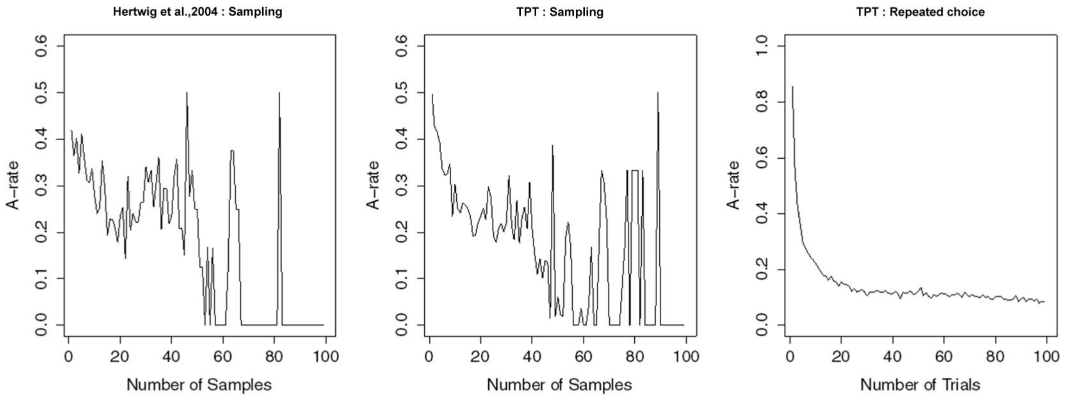

Figure 1. The A-rate (alternation rate between options) for the two sampling paradigm data sets and the first block of the repeated-choice data set analyzed by Gonzalez and Dutt (2011): Hertwig et al. (2004) and the Technion Prediction Tournament (TPT) data set (Erev et al., 2010). The A-rate is aggregated first within individuals and then over individuals at each sample size and choice number. The figures all reveal a gradual decay in the A-rate over time, but many of the high A-rate participants have dropped out of the sample in the right-hand side of the sampling data sets. Note that the figures replot the data in Gonzalez and Dutt's Figures 2A, 3A, and 3B (human data).

(Hau, Pleskac, Kiefer, & Hertwig, 2008; Hertwig & Pleskac, 2010; Ungemach, Chater, & Stewart, 2009). Note that calculating an aggregated A-rate over trials in the repeated-choice paradigm is not problematic because here the experimenter keeps the number of trials constant across individuals (see, e.g., Erev et al., 2010).

Figure 3A, however, still plots an aggregated A-rate. To reveal individual trends in exploration, we calculated the A-rate for the first 25% and last 25% of the sampling sequence for each person, alongside the same intervals for the repeated-choice data. Figure 3B plots the initial and final A-rates. If indeed exploration even-

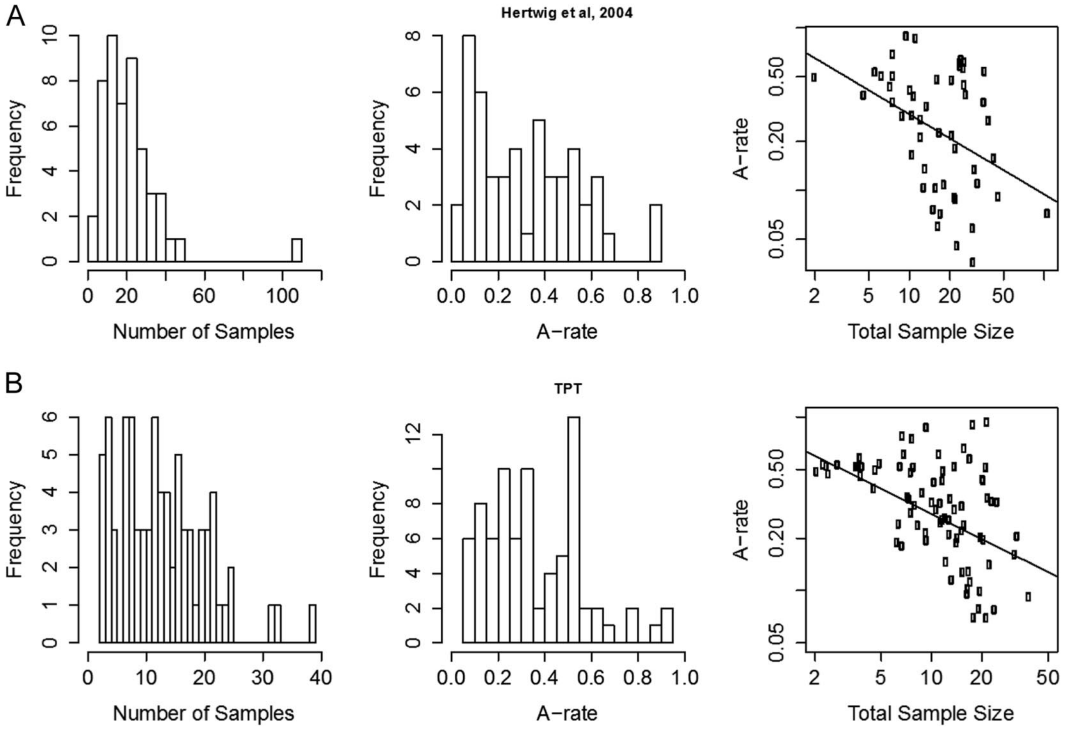

Figure 2. Histograms and correlations for total number of samples taken and the A-rate for the two sampling paradigm data sets analyzed by Gonzalez and Dutt (2011): Hertwig et al. (2004) and the Technion Prediction Tournament (TPT) data set (Erev et al., 2010). Lines in figures on the right-hand side represent the best fitting regression lines between log of total number of samples and the log of the overall A-rate.

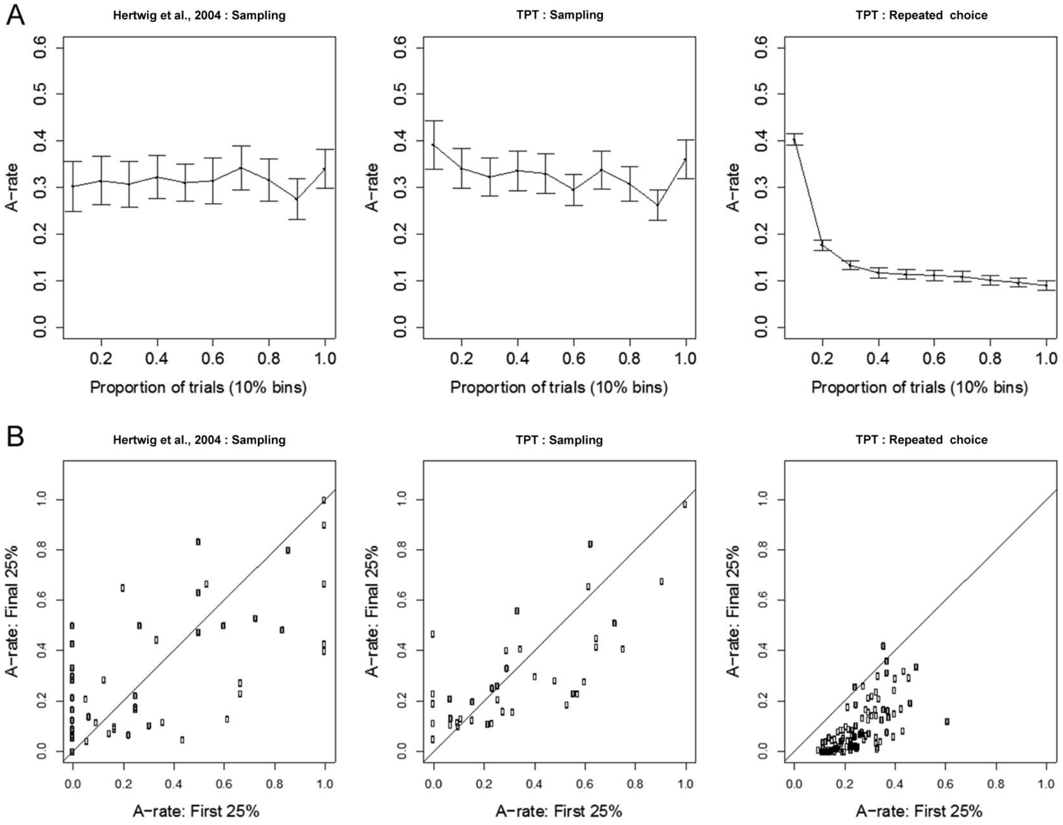

Figure 3. The mean normalized A-rate (alternation rate between options) for the two sampling paradigm data sets and the repeated-choice data analyzed by Gonzalez and Dutt (2011): Hertwig et al. (2004) and the Technion Prediction Tournament (TPT) data set (Erev et al., 2010). (A) The A-rate normalized by proportion of the sampling interval in 10% bins. Error bars indicate standard error. (B) The A-rate for the first 25% and the last 25% of the normalized sampling sequences. Squares on the diagonals represent people whose initial and final A-rates are identical.

tually superseded exploitation in the sampling paradigm, all data points should cluster in the triangle below the diagonal. This, however, is not the case. Instead, a minority of 10% (Hertwig et al., 2004) and 5% of people (Erev et al., 2010) for the two sampling data sets, respectively, have constant initial and final A-rates. The remaining participants fall in about equal size classes: In the Hertwig et al. (2004) data set, 23 show a reduction in A-rate, whereas 22 show an increase in A-rate. In the TPT data set, 18 show a reduction, and 20 show an increase. Averaging across all participants (using normalized data) yields the constant A-rate plotted in Figure 3A; on an individual level, however, about an equal portion of people appear to increase versus decrease exploration in the sampling paradigm (Figure 3B). The repeated-choice data, however, demonstrate a clear transition from high to low A-rates from the first to the last 25% of repeated choices. Only two out of 100 participants do not show this trend.

Ensuing Problems for the IBL Model

These findings highlight the dangers of drawing inferences from the A-rate averaged over trials in the sampling paradigm to individual search behavior, and they matter for the IBL model in its current form. If one accepts the constant average A-rate calculated across the normalized sequence as a more veridical representation of the aggregate relative to the declining average A-rate, then at least the following problems arise: First, the inferred isomorphism in the A-rate across the sampling and the repeated-choice paradigm disappears, making both paradigms less similar than suggested by Gonzalez and Dutt's (2011) analysis. Second, the probability of retrieving instance *i* from memory, and its activation (see Equations 3 and 4 in Gonzalez & Dutt, 2011), cannot change as predicted for both the repeated-choice and sampling paradigms. Specifically, in the sampling paradigm, there would appear to be no effect of the magnitude of the blended value for each option (*V* in Equation 2 from Gonzalez & Dutt, 2011), because the probability of choosing a specific option does not change systematically over the sampling interval. If the blended value of one option grew larger than the other option—as it must if the final choice is to be predicted accurately—then this would lead to less exploration over time as the sampling of one choice is favored over the other. However, as noted above, the data from the sampling paradigm do not show a systematic reduction in exploration at the individual level. This is a notable difference from the choice behavior observed in the repeated-choice paradigm.

Third, the IBL model involves the adjustable parameter *pInertia*, which determines whether the choice made in the previous trial is repeated. If *pInertia* = 1, then the IBL model will always repeat the last choice, or in other words, it will predict no exploration. Gonzalez and Dutt (2011) concluded that "in both paradigms, the A-rate decreases over an increased number of samples or trials" (p. 538), and that "results suggest that in both paradigms, humans move gradually from the exploration of options to their exploitation using the same cognitive mechanisms for the sequential selection of alternatives" (p. 539). The *pInertia*, estimated from human data, however, does not support this conclusion: They widely diverge across paradigms, inconsistent with the conclusion that behavior in both paradigms is "equivalent . . . at the sequential process (A-rate) level" (p. 539).[1]

The fourth problem the IBL model faces is that in the sampling paradigm it fails to predict the inverse correlations between A-rate and total sample size (see Figure 2, right panels). In order to account for total sample size, the IBL model randomly draws a value from distributions fitted to the empirically observed sample sizes in the Hertwig et al. (2004) and TPT data sets (Erev et al., 2010; Gonzalez & Dutt, 2011, p. 527). Consequently, the IBL modeling uses sample size distributions that have no relationship to the equations that generate alternations. Ergo, according to the IBL model, there is no relationship between sample size and alternations.

Conclusion

The IBL model is a very commendable attempt to explain the behavior in two experimental paradigms using the same cognitive mechanisms. And, indeed, the sampling and the repeated-choice paradigm have been found to produce surprisingly similar choice patterns, leading to a similar kind of description–experience gap (see Hertwig & Erev, 2009). This similarity, however, may not generalize equally well to other cognitive dimensions. In contrast to Gonzalez and Dutt's (2011, p. 539) conclusion, there appears to be no general move from exploration of options to their exploitation in the human data of the sampling paradigm. Their conclusion results from making inferences about individual behavior using data aggregated over individual trials, when individuals are leaving the aggregation at different numbers of trials in a way that is systematically related to the behavior being measured. Dealing with similar issues, Estes and Maddox (2005) highlighted the "danger . . . that individual differences among subjects with respect to values of a model's parameters may cause averaging to produce distorted inferences about true patterns of individual performance and the cognitive processes underlying them" (p. 403).

But let us not throw the baby out with the bathwater. The IBL model predicts final decisions in the sampling paradigm as well as or better than any other proposed model. It can also predict choices in the repeated-choice paradigm better than other models. This is obviously an excellent model on the level of choice, and some of its building blocks (e.g., the activation mechanism) have been demonstrated to be instrumental in successfully modeling a wide range of behaviors (e.g., Anderson & Lebiere, 1998, 2003; Gonzalez, Best, Healy, Kole, & Bourne, 2011; Gonzalez & Lebiere, 2005). However, the model in its present form suggests a deep similarity in the relative balance of exploration and exploitation between consequential and nonconsequential search. Empirically, however, this similarity is not so apparent—the human data do not appear to show a general move from exploration to exploitation in the sampling paradigm data (see Figure 3). Consequently, the IBL model's suggestion of such a general move is perplexing.

[1] Across two calibration sets, the values diverge in opposite directions. In one set, the calibration resulted in *pInertia* = 0.22 and 0.48 for the sampling and repeated-choice paradigm, respectively. In the other calibration set, however, the order is reversed and *pInertia* = 0.63 and 0.09 in the sampling and repeated-choice paradigms, respectively. We do not know why these values vary so widely and in different directions—but to the extent that *pInertia* is correlated with A-rate, their divergence does not support the conclusion that behavior in both paradigms is "equivalent."

References

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

Anderson, J. R., & Lebiere, C. (2003). The Newell test for a theory of cognition. *Behavioral and Brain Sciences, 26,* 587–601. doi:10.1017/S0140525X0300013X

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

Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review, 12,* 403–408. doi:10.3758/BF03193784

Gonzalez, C., Best, B. J., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2011). A cognitive modeling account of simultaneous learning and fatigue effects. *Cognitive Systems Research, 12,* 19–32. doi:10.1016/j.cogsys.2010.06.004

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. *Psychological Review, 118,* 523–551. doi:10.1037/a0024558

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

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making, 21,* 493–518. doi:10.1002/bdm.598

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15,* 534–539. doi:10.1111/j.0956-7976.2004.00715.x

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences, 13,* 517–523. doi:10.1016/j.tics.2009.09.004

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition, 115,* 225–237. doi:10.1016/j.cognition.2009.12.009

Hills, T. T., & Hertwig, R. (2010). Information search in decisions from experience: Do our patterns of sampling foreshadow our decisions? *Psychological Science, 21,* 1787–1792. doi:10.1177/0956797610387443

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

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science, 20,* 473–479. doi:10.1111/j.1467-9280.2009.02319.x

Yechiam, E., Busemeyer, J. R., Stout, J. C., & Bechara, A. (2005). Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychological Science, 16,* 973–978. doi:10.1111/j.1467-9280.2005.01646.x

Received November 8, 2011
Revision received January 16, 2012
Accepted January 30, 2012 ■