ERIC M. EISENSTEIN and J. WESLEY HUTCHINSON*

In this article, the authors examine the costs and benefits of action-based learning (i.e., learning that occurs as a by-product of making repeated decisions with outcome feedback). The authors report the results of three experiments that investigate the effects of different decision goals on what is learned and how transferable that learning is across related decision tasks. Contrary to popular wisdom, compared with traditional learning, experiential learning is likely to be a risky proposition because it can be either accurate and efficient or errorful and biased.

Action-Based Learning: Goals and Attention in the Acquisition of Market Knowledge

Expertise derived from experience is widely regarded as valuable. Managers, market mavens, and applicants to MBA programs rightly view their years of experience as an important credential. Moreover, experience is viewed as intrinsically different from the types of "book learning" that occur in schools. However, experiential learning in the marketplace does not occur for its own sake or for the sake of achieving high evaluations from an instructor. Rather, learning frequently occurs as a by-product of repeatedly making decisions about concrete actions and then observing the outcomes. Thus, such learning is known as "action-based" learning. Importantly, although each decision has its own goal, these goals are typically short-term and unrelated to globally accurate learning, per se. For example, product managers analyze customers and competitors, choose actions, and observe outcomes. As a by-product, they form impressions about the effectiveness of their actions. Consumers shop for products, observe prices, make purchases, and assess their satisfaction. As a by-product, they form impressions about the relationship between product attributes and market value. In such everyday environments, most experiences afford many dimensions of potential learning and many forms of feedback. Some feedback is important for immediate goals and some is important for the development of general knowledge; however, there is no guarantee that these two types of feedback coincide. Thus, it is natural to suspect that real-world learning, despite its high content validity, may be less complete, less accurate, and more biased than learning that has an explicit goal of high factual accuracy for an entire domain.

In this article, we adapt and extend the experimental paradigms of cognitive psychology to investigate the effects of different decision goals on action-based learning and how transferable that learning is to new situations (e.g., Brehmer and Brehmer 1988; DeLosh, Busemeyer, and McDaniel 1997; Eisenstein 2003; Juslin, Olsson, and Olsson 2003; Medin and Schaffer 1978). We demonstrate that action-based learning can be either accurate and efficient or errorful and biased and that decision makers who rely on action-based learning face substantial risks of missing significant environmental relationships, overgeneralizing knowledge, and exhibiting poor transfer of learning, even when task differences are small. Furthermore, we propose a new model-based explanation for why these effects are observed, which enables us to delineate the circumstances under which decision makers are most at risk.

ACTION-BASED LEARNING

The connection between decision making and learning is that the decision maker often must make predictions, based on available inputs, about the outcomes that will result from the actions that are available at the time of a decision. In some situations, the predictions required for decision making are consistent with maximizing the global accuracy of prediction. For example, an investor who is interested in real estate might review the specifications of a particular property and attempt to estimate both its current market value and its value at some time in the future (e.g., in dollars). This is a "numerical prediction task" (Brehmer and Brehmer 1988; Tversky, Sattath, and Slovic 1988). It is needed in this situation because the action space of the decision maker is continuous (i.e., making purchase offers), and

*Eric M. Eisenstein is an assistant professor, Johnson School, Cornell University (e-mail: eric.eisenstein@johnson.cornell.edu). J. Wesley Hutchinson is Stephen J. Heyman Professor of Marketing, Wharton School, University of Pennsylvania. The authors thank Jonathan Baron, Eric Bradlow, Jerry Busemeyer, Teck-Hua Ho, Stephen Hoch, Barbara Mellers, Stijn van Osselaer, Vithala Rao, and the three anonymous *JMR* reviewers for their helpful comments on previous versions of this article. Portions of this research were part of a doctoral dissertation at the University of Pennsylvania. The Wharton School funded this research.

ISSN: 0022-2437 (print), 1547-7193 (electronic)

success is dependent on point estimates of market value. However, many (arguably most) actions are discrete. For example, a person shopping for a new home might review the same information as the real estate investor to decide whether a particular property is affordable. The discrete action is whether to consider the house further (i.e., contact the realtor, visit the property, and so forth). This is a "categorical prediction task" (Medin and Schaffer 1978; Tversky, Sattath, and Slovic 1988) rather than numerical prediction.[1] That is, the property is either above or below the maximum amount the shopper can afford. In this case, global accuracy is not needed. Rather, accurate separation of affordable and unaffordable homes is needed. In this research, we focus on categorical prediction tasks that are similar to this, for which categories are defined by "cutoff" values of a criterion variable that could also be numerically predicted.

To account for action-based learning in numerical and categorical prediction tasks, we propose a general model in which the market environment creates "decision tasks" that are defined in terms of possible actions. These tasks evoke different decision-maker goals; in turn, goals differentially affect attention to stimuli, and differences in attention determine what is learned. We further propose that in addition to the immediate decision goal, there is the related "learning goal" of reducing future error. The central assumption in our model is that the task defines what counts as an error. Numerical prediction tasks naturally evoke deviation-based measures of accuracy (e.g., the difference between the predicted and actual numerical value), whereas categorical prediction tasks evoke classification-based measures of accuracy (e.g., proportion correct [PC]). We propose that increasing levels of attention and cognitive effort are allocated to decisions when the risk of error is believed to be high (and especially when outcome feedback reveals that error was, in fact, high). The end result of these additional cognitive resources is to update knowledge and thus reduce future error (in ways that we discuss in more detail subsequently). However, this updating process is not uniform. It reflects the definitions of error and the learning goals that the decision task induces. In this sense, the learning is action based.

Returning to our previous example, a real estate investor engaged in a numerical prediction task is likely to attend to the magnitude of error in his or her price estimate. If the estimate were off by $1,000, the investor would probably be satisfied, but if it were off by $100,000, the investor would be likely to change his or her beliefs about what attributes drive market value. In contrast, a potential home buyer engaged in the category prediction task of determining affordability will attend mainly to classification errors. If the home buyer had a cutoff of $500,000 and were to judge a $550,000 home to be $450,000 and, therefore, affordable (i.e., a classification error), the home buyer would likely update his or her beliefs. However, if the home buyer were off by the same amount in the opposite direction and were to judge the home to be $650,000, the home buyer would be much less likely to update his or her beliefs because the home was correctly judged to be too expensive.

Learning Categorical Versus Continuous Responses

Historically, there has been a split between research on the learning of continuous and discrete functions of multiattribute stimuli (DeLosh, Busemeyer, and McDaniel 1997; Juslin, Olsson, and Olsson 2003). For example, psychological research on concept formation has typically investigated category learning using a binary classification task (frequently based on binary attributes) and has primarily focused on testing specific models (e.g., rule-based versus exemplar-based models; see Ashby and Maddox 2005; DeLosh, Busemeyer, and McDaniel 1997; Kruschke 1992; Medin and Schaffer 1978; Nosofsky 1984). In marketing, studies using the categorization paradigm have investigated consumer learning of brand or category membership on the basis of categorical cues (e.g., Huffman and Houston 1993; Hutchinson and Alba 1991; Van Osselaer and Alba 2000). In addition, some studies have used related learning models to study the development of brand-attribute associations (e.g., Van Osselaer and Janiszewski 2001). In general, however, categorization research in marketing has placed greater emphasis on category definition and structure, the categorization process, and additions and extensions to a category rather than on the learning of categories, per se (e.g., Alba and Hutchinson 1987; Cohen and Basu 1987; Moreau, Markman, and Lehmann 2001; Sujan 1985; Sujan and Bettman 1989).

The learning of functional relationships between continuous cues and a continuous criterion has been most extensively pursued in psychology by means of the multiple-cue probability learning (MCPL) paradigm. Research on MCPL has shown that when study participants are given sufficient trials and appropriate feedback, a wide variety of functional relationships between cues and outcomes can be learned but that some relationships are more difficult to learn than others (for a review, see Brehmer and Brehmer 1988; see also Camerer 1982; Mellers 1980). In particular, linear functions with positive coefficients and no interactions are the easiest to learn, whereas interactions are very difficult to learn under most circumstances. In marketing, several researchers have used the MCPL paradigm. Meyer (1987) demonstrates that consumers can quickly learn positive attribute–quality relationships in a multiattribute setting. West (1996) uses an agent learning task to demonstrate that veridical feedback eliminates the facilitative effects of initial preference similarity and that learners allocate more time to informative than to uninformative feedback. Hutchinson and Alba (1997) demonstrate a framing effect in a budget allocation task in which prior allocations and their sales results served as the stimuli and feedback. Hoch and Schkade (1996) use a managerial credit-forecasting and learning task to investigate the design of decision support systems.

Although there is a large volume of research on categorization and multiple-cue judgment, the research seldom overlaps, and the two experimental paradigms are seldom combined in the same study (for notable exceptions, see DeLosh, Busemeyer, and McDaniel 1997; Eisenstein 2003; Juslin, Olsson, and Olsson 2003).

[1]We adopt Tversky, Sattath, and Slovic's (1988) terminology and contrast numerical prediction with categorical prediction. Juslin, Olsson, and Olsson (2003) use the terms "categorization" and "multiple-cue judgment." We prefer the former because it emphasizes both the similarity (i.e., prediction) and the differences (i.e., continuous versus discrete responses) between the two tasks.

Judgment, Choice, and Learning

The distinction we make between numerical and categorical prediction is similar to the distinction that is often made between judgment and choice, respectively (e.g., Johnson and Russo 1984; Payne, Bettman, and Johnson 1993; Tversky, Sattath, and Slovic 1988); however, the judgment versus choice and prediction paradigms are quite different. The key difference is that prediction tasks have objectively correct responses, but judgment and choice tasks are typically based on subjectively determined preferences. In the absence of objective criteria, the literature on judgment and choice has devoted much attention to the problem of assessing the rationality of decision behavior (e.g., Hastie 2001; Mellers, Schwartz, and Cooke 1998; Payne, Bettman, and Johnson 1993; Shafir and LeBoeuf 2002). In contrast, research using the prediction paradigm assesses the correspondence between beliefs and the correct answers (e.g., Brehmer and Brehmer 1988; Camerer 1982; Medin and Schaffer 1978). Another major difference between judgment versus choice and prediction pertains to the research paradigm and emphasis. Research in the subjective-preference paradigm has typically examined search, acquisition, and recall of information or has tested abilities that are based on knowledge acquired through education and personal experience before data collection. These preference paradigms stand in contrast to those that focus on learning or the development of conceptual knowledge; such paradigms emphasize performance gains that apply to new and old informational inputs, provide objective feedback during a training period, and attempt to understand the process through which concepts are added to a person's general base of knowledge.

Despite these substantial differences between the domains, they are clearly similar when regarded as forms of multiattribute decision making. Tversky, Sattath, and Slovic (1988) explicitly note this similarity and show how their contingent weighting model can account for results from both preference-based and prediction tasks. Although the prediction data they use were collected without feedback and their model does not attempt to account for learning, the decision process they describe for numerical and categorical tasks is easily extended to address our results. We discuss this in greater detail subsequently.

ATTENTION-BAND MODEL OF ACTION-BASED LEARNING

As discussed previously, we propose that people are likely to allocate additional attention to their decisions when the risk of error is high, or when feedback confirms that a significant error has been made. However, numerical and categorical prediction tasks invoke different definitions of error, and therefore attentional allocation is likely to be different for each type of task. It is likely that learners who are given numerical prediction tasks will pursue the goal of learning to predict the criterion variable as accurately as possible for all stimuli (i.e., maximizing global accuracy). However, learners who are given a categorical prediction task have two possible strategies: First, they could adopt the same goal as for numerical prediction. That is, they could predict the criterion variable as accurately as possible and base their binary classification response on their implicit predictions of this continuous variable. Second, their strategy could be more action based. Consistent with our central thesis, the learning goal would be determined by the decision task, and error would be defined as incorrect classification. Knowledge about the continuous criterion variable would be acquired incidentally, if at all.

Action-based learning strategies naturally focus attention selectively on stimuli with true values near the cutoff. There are two mechanisms that could result in this focus of attention. The first mechanism is reactive. Learners may exert increased cognitive effort only when stimuli are misclassified (i.e., reacting to error). This strategy naturally leads to a focus of attention around the cutoff, regardless of whether the learner is aware of it, because stimuli with criterion values near the cutoff are the most likely to be misclassified. The second mechanism is proactive. Learners may consciously realize that stimuli near the cutoff are most relevant to their learning because they are the most difficult. Thus, they deliberately allocate increased attention to stimuli near the cutoff, both during decision making and when evaluating feedback (regardless of whether feedback was positive or negative). Although the proactive and reactive mechanisms differ qualitatively, they are not mutually exclusive, and both result in selective attention on stimuli near the cutoff.

There is a long tradition of learning assessment in multiattribute prediction tasks using linear regression models, which are referred to as "cue abstraction models" (for a review, see Brehmer and Brehmer 1988; see also Dawes and Corrigan 1974; Einhorn 1972; Hammond 1955; Meyer 1987; West 1996).[2] Such models provide revealed (versus self-reported) measures of beliefs about how stimulus attributes are related to the criterion, similar to partworths in conjoint analysis. We develop an extension of the cue abstraction model that is designed to represent the effect of selective attention. The key component of this model is the introduction of an attention band in the criterion variable that defines error and determines the learning goal for a given task.

The Attention-Band Model

Let Y_t be the true value of the continuous criterion variable provided as feedback for the stimulus $\mathbf{X}_t = (X_{1t}, X_{2t}, ...)$ at trial t. The learning goal is represented by a function, $G(Y_t)$, of the presented feedback that the learner uses to improve the current decision rule for making predictions. More specifically, it is assumed that the learner minimizes deviations from this goal rather than deviations from feedback. A general specification of $G(Y_t)$ is as follows:

$$(1) \qquad G(Y_t) = \begin{cases} Y_L, \text{ if } Y_t < L, \\ Y_t, \text{ if } L \leq Y_t \leq U, \\ Y_U, \text{ if } Y_t > U, \end{cases}$$

[2]Such models are often called "paramorphic" because the estimated regression weights are meant to capture abstract aspects of decision rules rather than the cognitive processes that implement them (Einhorn, Kleinmuntz, and Kleinmuntz 1979). That is, the model is linear, but it is not assumed that the process involves people doing arithmetic in their heads. Rather, people rely on some attributes more than others, and coefficients are robust indicators of relative importance.

where L and U are constants. This definition captures the idea that people focus their attention on a specific range of outcomes (i.e., [L, U]), which we call the "attention band."[3] Within this range, the goal is to learn to predict specific values, so $G(Y_t) = Y_t$. Outside this range, the goal is to learn to predict general categories for "high" and "low" values (represented by single values, Y_L and Y_U, for each category). The goal function is used to develop a response function, $R(\mathbf{X}_t)$.

Cue abstraction models specify a linear response function with the following general form:

$$R(\mathbf{X}_t) = \mathbf{X}_t \boldsymbol{\beta}_t + \varepsilon_t, \tag{2}$$

where $\boldsymbol{\beta}_t$ is a column vector of weights at time t and ε_t is stochastic error. The learner adjusts these weights over time to minimize the discrepancy between $R(\mathbf{X}_t)$ and $G(Y_t)$.

The attention-band model represents a family of learning strategies. For example, during the training phase of the experiment, if the learner is given a categorical prediction task with a cutoff of Y_c, then L and U are chosen close to Y_c (including the possibility that $L = U = Y_c$, which yields a binary function). If the learner is given a numerical prediction task during training, then L and U are chosen such that [L, U] includes the full range of expected stimuli. During the test phase of the experiment, if the task were categorical prediction with a cutoff of Y_c' (and Y_c' may be different from Y_c), the response to stimulus $\mathbf{X}_t$ would be "above" if $R(\mathbf{X}_t) \geq Y_c'$, and it would be "below" if $R(\mathbf{X}_t) < Y_c'$, where $R(\mathbf{X}_t)$ is computed with the final values of $\boldsymbol{\beta}_t$ from the training phase. Analogously, if the test task were numerical prediction, the response to stimulus $\mathbf{X}_t$ would simply be $R(\mathbf{X}_t)$. Because no feedback is given during the test phase of most learning experiments, the model assumes that the weights used at the end of training are applied during the test phase without further updating.[4]

In some important ways, the attention-band model is similar to the contingent weighting model that Tversky, Sattath, and Slovic (1998) propose. Their model does not consider the role of feedback in learning; rather, it focuses on judgments and predictions that are based on multiattribute inputs. As is the case with our attention-band model, the contingent weighting model proposes that the nature of the task influences the weights applied to each attribute value in determining a prediction. However Tversky, Sattath, and Slovic's model is driven by the compatibility principle, which holds that when the task is quantitative in nature (e.g., numerical prediction), quantitative (i.e., continuous) inputs are more heavily weighted than qualitative (i.e., categorical) inputs, and conversely, when the task is qualitative in nature (e.g., categorical prediction), qualitative inputs are more heavily weighted than quantitative inputs. Thus, similar to the attention-band model, predictions are task dependent. However, the mechanism that Tversky, Sattath, and Slovic propose is different from the proactive and reactive attention-band mechanisms we described previously. The attention-band model proposes that the learning goal is "compatible" with the decision goal of the task. However, the quantitative or qualitative nature of the inputs is not relevant for the attention-band model. The contingent weighting and attention-band models are not mutually exclusive; however, they do make different predictions about our experimental results.

HYPOTHESES

General Hypotheses

Several general hypotheses that are robust with respect to stimulus design can be derived from the attention-band model.

> H_1: *Selective Attention*: Participants given categorical prediction tasks will allocate increased attention to stimuli near their cutoff compared with those given numerical prediction tasks.

This hypothesis follows from the assumptions that the learning goal, $G(Y_t)$, defines error and that attention is selectively given to stimuli that have a high risk of error. Because $G(Y_t) = Y_t$ inside the attention band and is constant outside the attention band, the risk of error, as defined by the goal, is higher inside the attention band (i.e., participants need to learn only Y_L and Y_U for stimuli that are outside the attention band, but they must learn to predict Y_t for stimuli that are inside the attention band). Participants given categorical prediction tasks should have relatively narrow attention bands, whereas participants given numerical prediction tasks should have very wide attention bands, effectively allocating attention equally across all stimuli. For example, a manager who is developing a digital camera targeted at a launch price of $300 is likely to pay more attention to other cameras that are priced near $300 than to cameras priced at $600. This type of selective attention to stimuli differs fundamentally from the attention mechanisms of most other models, which conceive of attention as "dimension stretching" (i.e., the effect of attention to an attribute is to expand participants' ability to discriminate among levels of that attribute; e.g., Goldstone and Steyvers 2001; Medin and Schaffer 1978; Tversky, Sattath, and Slovic 1988). Virtually all traditional models include a scale parameter, or weight, for each attribute. In these models, attentional allocation translates more or less directly into the size of that parameter. In the attention-band model, the attention that is focused on stimuli within the band can affect attribute weights in various ways that depend on the stimulus design, which leads to our next hypothesis.

> H_2: *Localized Learning*: Attention around the cutoff will result in learning that reflects the local characteristics of the stimuli within the attention band.

Focusing attention around the cutoff should result in learning that is biased to the extent that stimuli within the attention band are not representative of the entire domain (i.e., by increasing the leverage of some stimuli relative to others). For example, in the prediction of the price of a digital camera, it is likely that the weights linking attributes and price are different for cameras in the $150 price range than for those in the $600 range. The stimulus design of our experiments enables us to use H_2 to make specific predic-

[3]Weber, Green, and Luce (1977) coined this term for a psychophysical model of loudness judgments, in which people were assumed to sample neural information more heavily within the band than outside of it. For a recent review of the relationship between attention and categorization, see Logan (2004).

[4]We tested this assumption with the data from Experiments 2 and 3 for participants who had matched training and test tasks. We found no statistically significant differences between the estimated coefficients at the end of training and test for any coefficient in either the prediction or the cutoff matched conditions. Details are available on request.

tions that differentiate the attention-band model from other models of attention and learning.

H_3: *Localized Accuracy*: Categorical prediction–trained participants will exhibit fewer errors for stimuli near their cutoff and more errors for stimuli away from their cutoff than will numerical prediction–trained participants.

According to the attention-band model, respondents attempt to minimize the discrepancy between $R(\mathbf{X}_t)$ and $G(Y_t)$, and thus we assume that $R(\mathbf{X}_t) \approx G(Y_t)$. Because $G(Y_t)$ is constant outside the attention band for categorical prediction (i.e., Y_L or Y_U), error (i.e., mean absolute deviation [MAD] = $|Y_t - R(\mathbf{X}_t)|$) should be higher outside the attention band, where $R(\mathbf{X}_t) \approx Y_L$ or $R(\mathbf{X}_t) \approx Y_U$, than inside the attention band, where $R(\mathbf{X}_t) \approx Y_t$. For numerical prediction participants, error should be more uniform across stimuli. Thus, if overall learning is approximately the same for both groups, categorical prediction participants will be superior for stimuli near their cutoffs and inferior for stimuli far from their cutoffs.

General Experimental Procedure

In the experiments reported here, we tested our hypotheses using tasks that are similar in most ways to those found in prior research; however, we developed a hybrid paradigm to investigate differences between categorical and numerical prediction tasks (this paradigm was contemporaneously developed by Eisenstein [2003] and Juslin, Olsson, and Olsson [2003]). We use this paradigm and its stimulus design to refine the general hypotheses further. A key aspect of the design is that the relationship between two attributes and the criterion variable changed across levels of a third attribute. This feature creates an environment that is similar to the effect of a company employing a two-tier marketing strategy to target segments that differ in their importance weights and willingness to pay, thus increasing the ecological validity of our experiments. Furthermore, this aspect of the design enables us to test for the effects of an attention band.

We constructed the stimuli in our experiments using a multiattribute model that contained two continuous and one categorical variable. The criterion variable, Y, was a linear function of three attributes, X_1, X_2, X_3, and their interactions:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \quad (3)$$
$$b_{13}X_1X_3 + b_{23}X_2X_3 + b_{12}X_1X_2 + b_{123}X_1X_2X_3,$$

where X_1 and X_2 take on values of {–2.5, –1.5, –.5, .5, 1.5, 2.5} and X_3 takes on values of {–1, 1}, yielding a total of 72 stimuli. The binary attribute, X_3, always had the largest coefficient. Except for some conditions of Experiment 3, the attribute weights (b_0, b_1, b_2, b_3, b_{13}, b_{23}, b_{12}, b_{123}) = (516, 32, 32, 188, 16, –16, 0, 0). As a result, all stimuli with $X_3 = -1$ had criterion values less than 500, and all stimuli with $X_3 = +1$ had criterion values greater than 500. We refer to these as "nonoverlapping" stimuli because there is a complete separation of values around 500, based on the value of X_3. This is an important feature of the design because some participants were given a categorical prediction task with a cutoff of 500. All experiments incorporated an interaction between X_3 and the two continuous attributes such that when X_3 was –1, the coefficient for X_1 was three times less than that of X_2 (i.e., 16 versus 48), and when X_3 was +1, the coefficient for X_1 was three times greater than that of X_2 (i.e., 48 versus 16). This property implies that if only stimuli greater than or less than 500 are considered, no interactions are present (and X_1 and X_2 have unequal weights), but overall, the weights on X_1 and X_2 are equal. This is also an important feature of the design because we use several manipulations to direct attention to stimuli less than 500. In addition, as we mentioned previously, this design abstracts important characteristics of many real-world markets by using the X_3 to create subcategories and by using the interactions to simulate different attribute–price relationships within those market tiers. The specific characteristics of our stimuli also enable us to create more sensitive tests of our hypotheses.

In all experiments, we randomly assigned participants to a numerical or a categorical prediction task. In the numerical prediction task, participants predicted price on the basis of the attributes; in the categorical prediction task, participants classified stimuli as either "above" or "below" a specific price (i.e., a cutoff value on the underlying criterion). We used two categorical prediction tasks, one in which participants judged whether the criterion value was greater than or less than 500 and one in which the cutoff was 325; we refer to these as the "500-cutoff" task and the "325-cutoff" task, respectively. Equal-length cover stories placed participants' learning in context and defined the ranges for the attributes and criterion. Attributes were always given intuitive meanings and were rescaled into realistic ranges, and the criterion was always the price of a good. The binary attribute was either "above average" or "excellent." Experiments consisted of a training phase followed by a test. During training, participants in all conditions were sequentially shown 72 stimuli about which they made a prediction. On each training trial, a red bar was displayed on the screen for several seconds. Then, the bar disappeared, and attribute values appeared on the screen, after which participants made task-appropriate responses. After each trial, participants were given outcome feedback about the accuracy of their response. For the numerical prediction task, feedback consisted of the true price of the good. For the cutoff tasks, feedback consisted of both the true category, "above" or "below" the cutoff price, and the true price (cf. Juslin et al. 2003). To facilitate learning, we required that participants record the task-appropriate feedback value for each stimulus. In cutoff conditions, this entailed recording the correct answer, either "above" or "below," next to their response. In the numerical prediction condition, participants recorded "low" or "high" in conjunction with whether their prediction was less than or greater than the actual price.

A test phase followed training in which all participants performed the numerical prediction task for the same stimuli used in training but without feedback. We provided participants with a short cover story and asked them to perform the numerical prediction task. This was a new task for those in cutoff conditions. We presented test stimuli in a random order that was minimally correlated with the training order.

Specific Hypotheses Generated from the Attention-Band Model

The design enables us to perform a sensitive test of the localized-learning hypothesis (H_2). We expected that participants who were given a categorical prediction task with a low (325) cutoff would define error as an incorrect classification and therefore would allocate most of their attention to stimuli less than 500 (i.e., the 36 stimuli for which $X_3 = -1$, which will contain the greatest number of misclassifications). As a result, compared with participants who were given a numerical prediction task, participants who were given a categorical prediction task should overweight X_2 and underweight X_1. In addition, they should underweight the binary attribute because it accounts for no variance among stimuli less than 500. The nonoverlapping property of our stimuli implies that participants who were given a categorical prediction task with a cutoff of 500 could achieve perfect performance based on the binary attribute (X_3) alone. Thus, we expected that participants who were given this task would also define error as an incorrect classification, and this would lead them to overweight the binary attribute and underweight the continuous attributes.

To confirm our qualitative predictions, we applied the attention-band model to the stimuli used in our experiments. Specifically, we estimated $R(\mathbf{X}_t)$ by regressing multiple specifications of $G(Y_t)$ onto $\mathbf{X}_t$, which represents the maximum learning possible given a particular learning goal and a least squares objective function. We used several specifications of $G(Y_t)$ to represent the numerical prediction and cutoff tasks, as the attention-band model prescribes.[5] The results of these model estimations were consistent with our qualitative reasoning and enabled us to generate, *ex ante*, a robust set of predictions about how learning will be affected by action-based tasks if participants use strategies that are consistent with the attention-band model. Thus, H_{2a} and H_{2b} are sensitive tests of H_2:

- H_{2a}: For participants given a 325-cutoff task, there will be an asymmetry in the importance weights for the two continuous attributes, such that $\beta_1 < \beta_2$. Participants given the other tasks will not exhibit this asymmetry. We test this using the difference in weights, $\beta_2 - \beta_1$.
- H_{2b}: The weighting of the binary attribute relative to the two continuous attributes will be least for the 325-cutoff task, the greatest for the 500-cutoff task, and intermediate for the numerical prediction task. We test this using the ratio $\gamma = \beta_3/(\beta_1 + \beta_2)$.

Note that these predictions differ from those of other plausible models that explain task and goal effects by shifts in importance weights. For example, as we noted previously, the contingent weighting model of preference and choice (Tversky, Sattath, and Slovic 1988) can be adapted to our task. The compatibility principle predicts that participants assigned to either cutoff task should give greater weight to the binary attribute than those assigned to the prediction task, which implies that γ should be greatest for the numerical prediction task and smaller and equal for the two categorical prediction tasks. In contrast, the attention-band model predicts that participants given the 325-cutoff task will underweight (not overweight) the binary attribute compared with participants given the prediction task.

EXPERIMENT 1

To test H_2 and H_3, we used the three tasks discussed previously during the training phase (i.e., the 325-cutoff, 500-cutoff, and numerical prediction tasks), and all participants performed the numerical prediction task in the test phase. We test H_1 in Experiments 2 and 3.

Method

Participants. Forty-six students enrolled in a consumer behavior course participated for partial course credit and for an opportunity to win a \$100 prize. We removed five participants from the analysis for failing to follow directions.

Procedure. We used the general procedure described previously with the following specifications: The criterion values to be learned were the prices of houses (Y), which were based on three attributes: lot size in acres (X_1), house size in square feet (X_2), and school district quality (X_3). We presented stimuli to groups of participants using a computer projector. Therefore, participants saw all stimuli for the same amount of time and in the same random order. During training, participants had 10 seconds to respond to a presented stimulus by writing their predictions on a sheet of paper. Those assigned to categorical prediction tasks circled the word "above" or "below"; those assigned to the numerical prediction condition wrote their best estimate of the house's value. After 10 seconds, the correct numerical answer (i.e., the true price) appeared on the screen for 10 seconds, and participants in the cutoff conditions were also shown the correct task-appropriate category (i.e., above or below the cutoff). In the test phase, all participants predicted prices and had 12 seconds to respond. We again collected data on sheets of paper.

Results

Localized learning. We used ordinary least squares (OLS) regression to estimate decision weights for each participant by regressing predicted prices from the test phase onto the three attributes and their interactions. Mean values for these estimated weights appear in Table 1. H_{2a} predicts that the 325-cutoff participants will place greater weight on X_2 than on X_1, which we tested using the within-subjects difference $\beta_2 - \beta_1$ as the dependent measure. As we predicted, there was a difference for the 325-cutoff condition ($t(13) = -2.21$, $p = .04$; 64% of participants) but no differences in the other conditions (all t's < .75).[6] As H_{2b} predicts, $\gamma = \beta_3/(\beta_1 + \beta_2)$ was the least for the 325-cutoff task and the greatest for the 500-cutoff task, which we tested by a linear contrast ($F(1, 38) = 9.80$, mean square error [MSE] = 1.47, $p = .003$).[7] As we discussed previously, this result violates Tversky, Sattath, and Slovic's (1988) compatibility principle.

[5]To represent categorical prediction, we estimated the model using $G(Y_t)$ as defined in Equation 1, with attention-band widths ranging from zero (i.e., a binary function separating "above" from "below") to a band that included approximately half of the training stimuli. Numerical prediction tasks were represented by a full-range band that included all 72 training stimuli. Technical details are available on request. In these regressions, we defined Y_L to be the mean of all stimuli with criterion values less than L, and we defined Y_U to equal the mean of all stimuli with criterion values greater than U. The choice of the mean is plausible because there is substantial evidence that people can automatically and intuitively estimate the mean of a series of numbers (Levin 1974).

[6]Throughout this article, we report the percentage of participants who directionally exhibited the tested behavior as protection against potential aggregation biases due to unobserved heterogeneity (see Hutchinson, Kamakura, and Lynch 2000).

[7]One commonality across conditions (and all other experiments) is that the estimated weights for interactions were close to zero, which is consistent with prior research (e.g., Brehmer and Brehmer 1988; Camerer 1982; Mellers 1980).

Table 1

RESULTS FOR EXPERIMENTS 1–3

		Experiment 1[a]			*Experiment 2*						*Experiment 3*			
Measure or Parameter	*Normative Value*	*325-Cutoff (Manager)*	*500-Cutoff (Manager)*	*Numerical Prediction (Manager)*	*325-Cutoff (Manager)*	*500-Cutoff (Manager)*	*Numerical Prediction (Manager)*	*325-Cutoff (Consumer)*	*500-Cutoff (Consumer)*	*Numerical Prediction (Consumer)*	*500-Cutoff (Manager)*	*Numerical Prediction (Manager)*	*Overlap 500-Cutoff (Manager)*	*Overlap Numerical Prediction (Manager)*
N		14	13	14	9	9	8	8	8	10	20	18	18	17
Mean Regression Coefficients														
β_0	516	482	510	503	493	509	515	517	499	507	496	507	499	492
β_1	32	30	28	36	28	27	28	22	18	30	26	32	26	30
β_2	32	44	27	35	34	22	27	42	27	31	22	32	33	33
β_3	188[b]	123	173	153	147	164	177	116	148	174	177	155	97	83
β_{13}	16	7	4	7	4	1	2	7	0	6	3	4	2	3
β_{23}	–16	6	0	–4	5	–1	0	–3	0	–4	–1	1	0	2
β_{12}	0	2	2	–1	0	1	0	3	0	0	3	–1	2	0
β_{123}	0	1	2	0	0	0	–1	3	2	0	1	–1	1	1
Hypothesis Metrics														
$\beta_2 - \beta_1$	0	14	–1	–1	6	–5	–1	20	9	0	–4	0	7	3
$\gamma = \beta_3/(\beta_1 + \beta_2)$	≈3	1.7	3.2	2.2	2.4	3.4	3.2	1.8	3.3	2.9	3.7	2.4	1.6	1.3
Latency Measures														
Decision time (training)					4.0	3.5	8.5	3.9	3.4	7.7	3.2	8.0	3.9	6.8
Feedback time (training)					2.8	3.2	3.5	4.1	3.1	3.9	2.8	4.2	3.0	3.3

[a]In thousands of dollars for Experiment 1 and in dollars for Experiments 2 and 3.

[b]β_3 = 110 in the overlap conditions of Experiment 3.

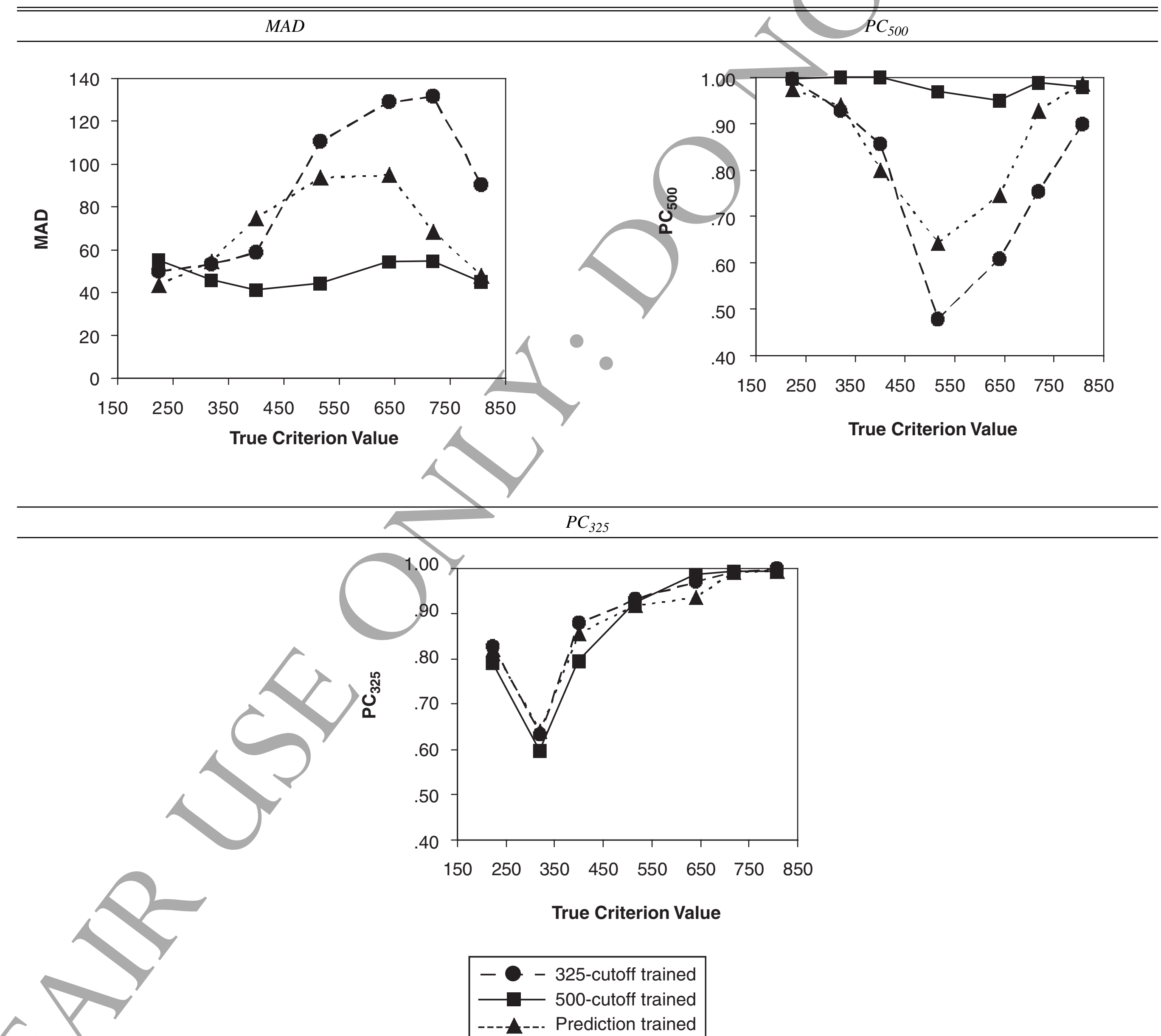
Localized accuracy. For each participant, we computed the MAD of the test phase numerical prediction from the true value, the PC for the 325-cutoff task (PC_{325}), and the PC for the 500-cutoff task (PC_{500}). We used the categorical predictions implied by the observed numerical responses to compute PC_{325} and PC_{500} for each participant. Thus, regardless of training task, we computed each measure of accuracy for each participant.

H_3 predicts increased accuracy around the cutoff for participants in the cutoff conditions. To visualize the relevant results, Figure 1 displays each performance measure as a function of the criterion for each condition. To simplify these plots, we grouped stimuli into seven bins according to criterion values: less than or equal to 275, 276–370, 371–425, 426–600, 601–675, 676–760, and greater than 760. These boundaries made the number of stimuli in each bin as equal as possible. Figure 1 (MAD panel) reveals that participants in the cutoff conditions were more accurate (i.e., MAD was lower) near their respective cutoffs than far from their respective cutoffs and that, in contrast, prediction condition participants were most accurate at the extremes.

We statistically tested the increase in accuracy around the cutoff by creating a problem-level effects-coded variable, BAND. For each cutoff condition, BAND was positive for all stimuli in the three bins in Figure 1 that are centered on the cutoff category and was negative for stimuli in the other

Figure 1

EXPERIMENT 1 MAD, PC_{500}, AND PC_{325} ACCURACY BY TRUE PRICE



Notes: The MAD estimates are the mean of the median within-subjects accuracy within each bin, and PC measures are the mean of the mean within-subjects accuracy within each bin.

bins. We chose the magnitudes of these values so that the mean across stimuli would be 0 and the spacing between the positive and negative codes would be equal to 1 (exact values depended on the number of stimuli in the three bins, which differed between conditions). BAND was 0 for the numerical prediction condition stimuli. We included dummy variables for participants and stimuli to control statistically for any general effects of ability or problem difficulty, and we used stimulus order to control for any effect of fatigue. We pooled data across conditions, and we regressed MAD onto BAND and the control variables. We summarize the results in Table 2. The significant, negative coefficient of BAND indicates that cutoff-condition participants had increased accuracy (i.e., lower MAD) near their cutoffs, as we predicted in H_3.

Consistent with the results for MAD, Figure 1 (PC_{500} panel) shows that the 500-cutoff participants performed better than those in other conditions in terms of PC_{500} around the 500 cutoff. Whereas the 500-cutoff participants (PC_{500} = .97) made essentially no errors in classification around their cutoff, the numerical prediction participants (PC_{500} = .72) were substantially less accurate ($F(1, 38)$ = 12.5, MSE = .03, p = .001). In contrast, the 325-cutoff participants were not more accurate around their cutoff ($F < 1$).

Discussion

The results of Experiment 1 support the attention-band model predictions of localized learning (H_2) and localized accuracy (H_3). The costs and benefits of selective attention were apparent. The enormous advantage in PC_{500} accuracy for 500-cutoff-condition participants (Figure 1, PC_{500} panel) clearly shows "return on investment" for attentional allocation near their cutoff. The extreme difference in PC_{500} between 500-cutoff participants and numerical prediction participants suggests that the former learned that the binary attribute (X_3) perfectly predicted whether price was greater than or less than 500, but the latter did not. The differences in γ (as we defined it in H_{2b}) reinforce this conclusion. Furthermore, 500-cutoff participants were more accurate than numerical prediction participants in overall MAD and in MAD around the 500 cutoff. In contrast, whereas the cost of selective attention for the 325-cutoff participants is evident in the large values of MAD when stimulus values were far from their cutoff, the benefits in classification performance were much less for this group (compare MAD and PC_{325} in Figure 1). Although the estimated importance weights for the 325-cutoff condition were consistent with the attention-band model, these differences were not sufficient to improve classification accuracy relative to the other conditions. We suspect that this is due to the greater difficulty of the 325-cutoff task relative to the 500-cutoff task. Finally, the results violate Tversky, Sattath, and Slovic's (1988) compatibility hypothesis. Overall, the results of Experiment 1 suggest that the costs and benefits of action-based learning are strongly dependent on the decision task (i.e., the location of the cutoff) and the empirical relationships among market variables.

EXPERIMENT 2

Overall, the results of Experiment 1 were consistent with our predictions. In Experiment 2, we focus on selective attention (H_1). Therefore, we introduced two substantive changes to the design. First, to track the allocation of attention more carefully, the experiment was self-paced, and we measured response latencies at each stage. H_1 predicts that in cutoff conditions, latencies will be longer for stimuli with prices near the cutoff than for those with prices far from the cutoff. Furthermore, self-pacing is more ecologically valid. Second, we manipulated the role that participants adopted to provide an additional test of attentional shifts. In Experiment 1, all participants adopted the role of a real estate manager. Given this perspective, there was no reason for them to focus attention on stimuli that were either above or below the cutoff in categorical prediction tasks. In Experiment 2, half of the participants adopted the role of a manager, and half adopted the role of a consumer. In cutoff conditions, we described the consumer role as an upper limit on spending. We expected this upper limit to shift attention downward toward products that might ultimately be purchased, which should increase the overweighting of X_2 relative to X_1. We did not expect role to have an effect for the numerical prediction condition.

H_4: *Role*: The consumer role will shift attention in the categorical prediction conditions toward stimuli with criterion val-

Table 2

LOCAL ACCURACY AND INCREASED ATTENTIONAL ALLOCATION AROUND THE CUTOFF

	Measure	*BAND*	*C-ERROR*	*P-ERROR*
Experiment 1	MAD	–35.47* (5.51)		
Experiment 2	Feedback time(s): no error	.52* (.10)		
	Feedback time(s): error	.35* (.10)	1.37* (.17)	.0060* (.0009)
	Decision time(s)	.62* (.14)		
	MAD	–14.52* (2.64)		
Experiment 3	Feedback time(s): no error	.51* (.10)		
	Feedback time(s): error	.31* (.09)	2.02* (.12)	.0058* (.0005)
	Decision time(s)	.61* (.20)		
	MAD	–12.28* (–3.25)		

*p < .001.

Notes: The standard error of each coefficient appears in parentheses. The following are explanations of items we controlled for: Problem ID = a dummy variable for each problem, participant = a dummy variable for each participant, and fatigue = a linear term that represents the ordinal position in which the problem was presented during training.

ues below the cutoff; therefore, the asymmetry in importance weights for the two continuous attributes (i.e., $\beta_2 - \beta_1$) will become more extreme.

Method

Participants. Sixty students participated for partial course credit and for a chance to win \$100. All participants were enrolled in a consumer behavior course. None had previously participated in other similar experiments. We removed eight participants from the analysis for failing to follow directions or for failing to learn during training.[8]

Procedure. We used the general procedure discussed previously with the following specifications. Participants were randomly assigned to one of the six experimental conditions: 3 (task: 325 cutoff, 500 cutoff, and numerical prediction) × 2 (role: manager and consumer). To make the task more relevant for the participants and to demonstrate the generality of the results obtained in Experiment 1, we changed the product category from houses to digital cameras. The cover story for consumers was screening for affordable products. The cover story for managers was screening for an appropriate sales channel under the assumption that cameras priced above the cutoff would be sold in a different channel than those below the cutoff. The criterion values to be learned were the prices of the cameras (Y). The presented values of the three attributes were identical to those in Experiment 1, but we labeled them as follows: hours of battery life (X_1), kilo-pixels of resolution (X_2), and optical zoom quality (X_3). We presented stimuli in six randomized orders.

The training procedure was the same as that in Experiment 1, except that participants could self-pace all aspects of training and test. Stimulus presentation and data collection were entirely computer based and individually administered. During the test phase, participants predicted the values of 80 digital cameras. The first 72 stimuli were the same as those we presented during the training phase. These stimuli were followed by an additional eight new stimuli.

Results

Selective attention. The latency times we collected in Experiment 2 represent a direct measure of attentional allocation. There are two latencies of particular interest: "Decision time" is the time between the presentation of the stimulus and the participant's response during training, and "feedback time" is the time spent examining feedback. The average of these latencies appear in Table 1.[9] Overall, decision time in the numerical prediction condition was longer than that in the cutoff conditions ($F(1, 46) = 82.1$, $MSE = 2.69$, $p < .0001$). There were no effects of task or role on feedback time (omnibus $F < 1$; see Table 1).

H_1 predicts that participants assigned to categorical prediction conditions will allocate increased attention to stimuli around the cutoff. We tested this by regressing decision time and feedback time onto BAND and the control variables that we used in Experiment 1. The coefficient for BAND was positive and statistically significant for both latency measures, in support of H_1 (see Table 2). However, as we discussed previously, there are potentially both proactive and reactive mechanisms that the attention-band model proposed that could cause this effect. The reactive mechanism postulates that attention-band effects are a consequence of learners allocating increased attention when they make an error (which naturally occurs more frequently around the cutoff), whereas the proactive mechanism postulates that there is increased attentional allocation regardless of error. The reactive mechanism can operate only for feedback times because participants do not know their error during decision making. Thus, the significant coefficient of BAND for decision times provides evidence for the proactive mechanism. To differentiate these two psychological mechanisms further, we added two variables to the regression analysis of feedback times, C-ERROR and P-ERROR (classification error and prediction error, respectively). C-ERROR was a stimulus-level dummy-coded variable that took on the value of 1 when cutoff participants incorrectly classified the stimulus and 0 otherwise (i.e., when categorical prediction participants correctly classified the stimulus and for all latencies in numerical prediction conditions). P-ERROR was the absolute deviation of numerical prediction for each trial (and 0 for all categorical prediction participants). Consistent with the reactive mechanism, C-ERROR was positive, statistically significant, and relatively large (1.37 seconds; see Table 2). Similarly, participants in numerical prediction conditions attended to larger errors for a longer amount of time, as is shown by the significant, positive coefficient of P-ERROR (see Table 2). This coefficient implies that an absolute deviation of 227 resulted in an increased latency approximately the same as that for a misclassified stimulus (i.e., 1.37 seconds). The significant coefficient of BAND demonstrates that cutoff participants allocated .35 seconds of additional feedback time to problems inside the band versus outside the band, even after we controlled for the effects of C-ERROR, supporting the presence of both proactive and reactive mechanisms.

Localized learning. As in Experiment 1, we used OLS regression to estimate decision weights for each participant. Mean values for these estimated weights appear in Table 1. As in Experiment 1, we tested H_{2a} using the within-subjects difference $\beta_2 - \beta_1$. Replicating the results of Experiment 1, we found a difference in weight in the 325-cutoff condition ($t(16) = -2.39$, $p = .03$; 65% of participants) but no difference in the numerical prediction condition or in the 500-cutoff condition (all t's < .50). As H_{2b} predicts, γ was the least for the 325-cutoff task and the greatest for the 500-cutoff task, which we tested using a linear contrast ($F(1, 46) = 6.09$, $MSE = 2.23$, $p = .02$). As we discussed previously, this violates the compatibility principle (Tversky, Sattath, and Slovic 1988). Furthermore, as H_4 predicts,

[8]The criterion for being classified as a "nonlearner" necessarily depended on the experimental cell. For participants in numerical prediction conditions, we used the criterion of an R-square between the training estimates and the actual prices that was less than .20. For the cutoff conditions, we used the criterion of a failure to achieve a final training accuracy of at least 70%.

[9]To reduce outliers (which are common in latency data), we computed medians for each participant before we computed means. Furthermore, throughout our analyses, we removed extreme latency times for each participant (following, e.g., West 1996). Total deleted data were 1.9% of observations, and the average was 1.4 latencies per participant. All presented conclusions are robust to the criterion that we used to define an outlier latency.

the role manipulation increased the asymmetry in weights for the two continuous attributes in both cutoff conditions, as we show by contrasting the difference $\beta_2 - \beta_1$ for the cutoff consumer and manager roles (see H_5 and Table 1; $F(1, 46) = 5.72$, $p = .02$).

Localized accuracy. As in Experiment 1, we used OLS regression to estimate the coefficient of BAND, which represents the difference in accuracy for stimuli inside versus outside the attention band for the categorical prediction conditions (see Table 2). The significant, negative coefficient of BAND indicates that participants in categorical prediction conditions displayed increased accuracy near their cutoffs, as H_3 predicts. As in Experiment 1, the 500-cutoff participants performed much better than those in other conditions in terms of PC_{500} accuracy around the 500 cutoff. Whereas the 500-cutoff participants (PC_{500} = .98) made essentially no errors in classification around their cutoff, the numerical prediction participants (PC_{500} = .86) were much less accurate for these stimuli ($F(1, 46) = 6.82$, MSE = .02, $p = .01$). As in Experiment 1, the participants in the 325-cutoff condition were not more accurate in PC_{325} around their cutoff ($F < 1$).

Discussion

Overall, these results replicate and extend the results of Experiment 1. In particular, the latency data provide direct support for H_1. Although participants assigned to categorical prediction tasks spent far less time deciding than did participants in the numerical prediction condition overall, they spent more time on stimuli with criterion values near their cutoff than on stimuli with values far from their cutoff, which is what would be expected from an attention-band model (for similar results using a different type of learning task, see Maddox, Ashby, and Gottlob 1998; see also West 1996). Further analysis confirmed that the additional time allocated to feedback within the attention band was a result of both proactive and reactive mechanisms.

Replicating the results of Experiment 1, we found that 325-cutoff participants exhibited the predicted asymmetry in decision weights for the continuous attributes (H_2). As we predicted, this asymmetry was more extreme for consumer than for manager cutoff roles (H_4). Furthermore, both cutoff conditions demonstrated localized accuracy (H_3). These results violate Tversky, Sattath, and Slovic's (1988) compatibility principle.

As in Experiment 1, there were clear costs in the 325-cutoff condition in terms of transfer of learning to the other tasks. However, 325-cutoff participants were equal to numerical prediction participants in accuracy around their cutoff (for MAD and PC_{325}) and took less time to achieve this level of performance. In contrast, the 500-cutoff participants were faster than numerical prediction participants, equally accurate in overall MAD, and substantially more accurate in PC_{500}, especially for the most difficult stimuli. Comparing the results of Experiments 1 and 2, we find no evidence that self-pacing damaged the learning process in any major way, because we replicated all results.

EXPERIMENT 3

The results of Experiments 1 and 2 were consistent with our general predictions about action-based learning and with the specific predictions of the attention-band model. Somewhat surprisingly, the 500-cutoff-trained participants in both experiments consistently performed as well as the numerical prediction–trained participants in terms of MAD and were superior in PC_{500} accuracy. Furthermore, this good performance was not the result of categorical learning per se (as the compatibility principle predicts), because we did not observe it for the 325-cutoff group. However, in the previous experiments, X_3 was a perfect predictor of the price being greater than or less than 500. One plausible hypothesis is that the superior classification performance of the 500-cutoff-trained group is due to the use of a simple verbal rule. This conjecture is consistent with the experimental and neuropsychological evidence that Ashby and Gott (1988) review (see also Nosofsky, Palmeri, and McKinley 1994). To test this conjecture, in Experiment 3, we presented half of the participants with a stimulus set in which X_3 was a good, but not a perfect, predictor of the price being greater than or less than 500, eliminating the simple verbal rule for classification. Thus, in half of the conditions in Experiment 3, b_3 was reduced to 110 (and all other coefficients were unchanged). As a result, in the 500-cutoff task, only 83% correct classifications could be achieved solely on the basis of X_3. We refer to these as "overlapping" stimuli because both levels of X_3 have some stimuli greater than and less than 500. As in previous experiments, we applied the attention-band model to the overlapping stimuli to generate specific hypotheses (details are available on request). The results of these regressions revealed that the near-perfect performance of the 500-cutoff condition should be reduced to a level that is comparable to that of the numerical prediction condition because of more errors near the cutoff.

> H_5: *Simple Rule*: The shift to the overlapping stimulus set should degrade the performance of the 500-cutoff condition more than the numerical prediction condition.

In Experiments 1 and 2, we tested all participants on the numerical prediction task. Although we found strong evidence for the attention-band model, it could be argued that the results were artifacts of the change during the test phase or the recoding of the numerical prediction data into categorical predictions. A plausible hypothesis might be that the numerical prediction group learned that the categorical attribute was a perfect predictor of classification greater than or less than 500, but without being cued during the test phase, they were unable to recall this helpful piece of knowledge when making their predictions. Use of the 500-cutoff task during the test phase should explicitly cue the importance of the simple verbal rule (and the significance of the 500 boundary in general) and should allow recall of any information that participants stored in their memory but did not retrieve in previous experiments. Distinguishing between encoding and retrieval mechanisms has important theoretical and pragmatic implications. Theoretically, we posit that selective attention at encoding determines what is learned, so a retrieval-based explanation undermines our model. Pragmatically, the costs and benefits of all learning strategies are reduced if they can be subsequently undone by providing or denying appropriate retrieval cues.

> H_6: *Encoding*: The effects of the attention band operate at the time of learning, not during recall. Therefore, explicit cuing

will not improve numerical prediction–trained participants' performance.

Method

Participants. A total of 154 students participated for partial course credit. None had previously participated in similar experiments. We removed 12 participants from the analysis for failing to follow directions or for failing to learn during the training phase, using the same criteria as in Experiment 2.

Procedure. We used the stimuli and procedures of Experiment 2 with the following changes: We randomly assigned participants to one of the eight experimental conditions: 2 (training task: 500 cutoff or numerical prediction) × 2 (test task: 500 cutoff or numerical prediction) × 2 (overlap: nonoverlapping or overlapping stimuli). The nonoverlapping stimuli and digital camera cover stories were identical to those in Experiment 2 with the exception of cover-story changes that were necessary to allow for the cutoff-test task. We described the overlapping stimuli previously. The change in the coefficient of X_3 from 188 to 110 had the effect of making a total of 10 exceptions (out of 72) to the simple verbal rule based on the value of X_3. All other aspects of the stimuli and procedure were the same as those in Experiment 2 except for the instructions for the 500-cutoff-test task.

Results

Selective attention. We computed median latency across trials and within-subjects for both decision time and feedback time, as we did in Experiment 2 (see Table 1). Replicating the results of Experiment 2, we found that the median decision time for numerical prediction–trained participants was nearly twice that of 500-cutoff-trained participants ($F(1, 139) = 80.1$, MSE = 8.80, $p < .0001$). Numerical prediction–trained participants also spent more time examining feedback than 500-cutoff participants ($F(1, 139) = 31.3$, MSE = 1.35, $p < .0001$).

As in Experiment 2, we tested selective attention (H_1) by regressing decision time and feedback time onto BAND and then onto BAND, C-ERROR, and P-ERROR (with both analyses including the previously defined control variables).[10] The coefficient for BAND was significant and positive for both latency measures, in support of H_1 (see Table 2). The second regression replicated the results of Experiment 2. C-ERROR was significant, positive, and relatively large (2.02 seconds; see Table 2), consistent with the reactive mechanism we discussed previously. Similarly, participants in numerical prediction conditions attended to larger errors for a longer amount of time, as is shown by the significant, positive coefficient of P-ERROR (see Table 2). This coefficient implies that an absolute deviation of 348 resulted in an increased latency approximately the same as that for a misclassified stimulus (i.e., 2.02 seconds). Finally, the significant, positive coefficient of BAND indicates that participants in categorical prediction conditions allocated .62 seconds of additional feedback time to problems inside the band versus outside the band, even after we controlled for the effects of C-ERROR; this finding supports the presence of both proactive and reactive mechanisms.

Localized learning. We tested one-half of the participants on the numerical prediction task. As in previous experiments, we used OLS regression to estimate weights for each participant by regressing predicted prices from the test phase onto the three attributes and their interactions. Mean values for these estimated coefficients appear in Table 1. H_{2a} is not applicable, because there is no 325-cutoff condition. H_{2b} predicts that γ will be greater for 500-cutoff participants than for prediction participants. An analysis of variance revealed a training task × overlap interaction ($F(1, 66) = 3.93$, MSE = .95, $p = .05$), which is due to the nonoverlap condition strongly showing this difference in γ, but the overlap condition was only directionally consistent.

Localized accuracy. As in previous experiments, we used regression to estimate the coefficient of BAND, which represents the difference in accuracy for stimuli inside versus outside the attention band for the cutoff condition (see Table 2). The significant, negative coefficient of BAND indicates that participants assigned to the categorical prediction task displayed increased accuracy near their cutoffs, as H_3 predicts. This result held for both the overlapping and nonoverlapping conditions. Replicating the results of previous experiments, we found that participants in the 500-cutoff nonoverlapping condition made essentially no errors in classification around their cutoff ($PC_{500} = .98$), but participants in the numerical prediction condition were much less accurate ($PC_{500} = .76$; $F(1, 135) = 56.3$, $p < .0001$). When we eliminated the simple verbal rule by using overlapping stimuli, the 500-cutoff-trained condition remained superior by approximately six percentage points ($F(1, 135) = 4.22$, $p = .04$), providing further support for H_3.

Simple rule and encoding. We computed PC_{500} for all participants. H_5 predicts that the overlapping stimuli will degrade the PC_{500} performance of the 500-cutoff condition more than the prediction condition. H_6 predicts that explicitly cuing the 500 boundary by testing on the 500-cutoff task will not improve numerical prediction–trained participants' performance. An analysis of variance on PC_{500} revealed effects of training task ($F(1, 135) = 45.8$, $p < .0001$), overlap ($F(1, 135) = 74.5$, $p < .0001$), and the training task × overlap interaction ($F(1, 135) = 5.21$, $p = .02$). As H_5 predicts, the interaction is due to the greater degradation in performance for the cutoff condition than for the numerical prediction condition. The effect of test task was marginal ($F(1, 135) = 3.02$, $p = .08$) and was not differentially beneficial to the numerical prediction condition (amounting to only a 1% increase in PC). This result indicates that unattended information is lost at the time of encoding, in support of H_6.

Discussion

Overall, these results replicate and extend the results of the previous experiments. Again, we found support for selective attention (H_1, both proactive and reactive mechanisms), localized learning (H_2), and localized accuracy (H_3). The removal of the simple verbal rule by using overlapping stimuli had the expected effect of degrading the performance of cutoff-trained participants more than numerical prediction–trained participants, in support of H_5.

[10]As in Experiment 2, we removed extreme latency times for each participant from the analysis (see fn. 9). Total deleted data were 2.1% of observations, and the average was 1.5 problems deleted per participant.

Nonetheless, the cutoff-trained participants remained superior in both overall accuracy and on the most difficult problems, even without a simple rule, which provides additional evidence for the beneficial effects of selective attention. Furthermore, as H_6 predicts, numerical prediction–trained participants were not helped significantly by explicit cuing of the importance of the 500 cutoff, supporting the conclusion that the results are driven by encoding effects during learning, not by subsequent problems with retrieval.

GENERAL DISCUSSION

In this research, we examined experiential learning in an environment in which multiattribute inputs were linked to a continuous criterion. This type of environment is commonplace in managerial and consumer decision making because it arises whenever a cutoff-based screening process is followed by evaluation of the considered options. For example, the environment we used in these experiments is analogous to consideration set formation, the screening of new product ideas for further development, and instruction about buying and selling prices. We made an important distinction between learning to estimate the criterion (numerical prediction) and learning to classify a stimulus correctly as above or below a specific cutoff value on that criterion (categorical prediction). This experimental paradigm resembles many real-world learning environments and differs from many traditional experimental paradigms, which usually emphasize either numerical or categorical prediction and provide unambiguous feedback.

At the most general level, our results suggest that learning is sensitive to the decision maker's goals during the learning process and therefore is not independent of the decisions that lead to learning. Relatively small changes in tasks and goals influenced how much attention participants paid to stimuli and to feedback, and these changes resulted in large differences in performance. In particular, we demonstrated that categorical prediction focuses attention on stimuli near decision cutoffs, which leads to localized learning and localized accuracy. In some situations (e.g., 325-cutoff conditions), overall accuracy suffered as a result of these attentional changes. In other situations (e.g., 500-cutoff conditions), overall accuracy was as good for people who were trained on numerical prediction, even though participants in the categorical prediction conditions invested much less time in learning. It is important to recall that the observed differences in learning between categorical and numerical prediction tasks support the general hypothesis that learning is action based. If people had the goal of extracting as much information from feedback as possible (an assumption often made in economic theory), they would adopt numerical prediction as the learning goal even for categorical prediction tasks. Although this strategy is more effortful, it has the advantage of preparing the learner for new decision tasks that might be encountered in the future. Our results show that there are advantages and disadvantages to action-based learning because accuracy depends on whether the learning goal directs attention away from or toward informative stimuli and whether the goal increases the likelihood of discovering important relationships among stimuli (e.g., a simple verbal rule). However, this dependence emphasizes the riskiness of relying on experiential (i.e., action-based) learning.

We used a price-learning task because this is an important and ubiquitous form of market knowledge. However, the presented results are applicable to any situation in which the relationship between multiattribute inputs and a continuous criterion must be learned, a frequent occurrence in decision-making situations in which there is a screening phase followed by an evaluative phase. For example, consumers who are comparison shopping for almost any durable good are at risk to the extent that they begin their search by screening products on the basis of one attribute and then shift to a different attribute or another range within the same attribute as their search progresses. Such shifts might occur because their needs change, because a salesperson successfully "upsells" them, or because new products enter the market. Similarly, salespeople themselves may be affected by upselling because if they habitually focus on the low end of the market, the shift to high-end products represents a change to a product line, which is not likely to have received much attention in the past. In general, the reported results suggest that managers and consumers should increase their use of objective analyses and decrease reliance on experience or intuition. This might be regarded as the standard view of academics and marketing research professionals; however, our results provide empirical evidence of its validity.

Attention-Band Models

To sharpen our qualitative predictions, we developed an attention-band model of decision making and learning. According to this model, the learner's strategy is represented by a learning goal that is a function of continuous outcome feedback. This goal requires higher levels of accuracy near the decision cutoff, and the region of higher accuracy is called the attention band. The attention-band model can be used in conjunction with almost any model of learning, including the cue abstraction and exemplar models that have been investigated extensively over the past 30 years.[11] Unlike traditional models of learning, the attention-band model assumes that learners do not strive for accuracy, per se, but for accuracy as defined by a specific learning goal. Therefore, attention is not represented directly as a weighting of attributes (i.e., "dimension stretching"), as it is in most learning models. Rather, attention is represented by the learning goal, insofar as more cognitive effort must be expended to learn about stimuli within the attention band than about other stimuli. Thus, it is assumed that high levels of attention are paid to all attributes of the stimuli in the attention band, but this distorts what is learned and induces shifts in the attribute weights used to predict the criterion variable. Our results support both a proactive mechanism, in which learners deliberately allocate increased attention to stimuli inside the band, and a reactive mechanism of attention, which is directed by error feedback.

[11]We successfully used the attention-band model with the original general-context model (Medin and Schaffer 1978) and an extension. The results were consistent with those we report herein, except that the original general-context model was strongly rejected as fitting these data. The extension performed as well as the cue abstraction model. A more technical working paper containing details is available on request.

Expertise

In the extensive literature on expertise, researchers have measured expert performance in various ways, and in general, they have found it to be surprisingly low. Much of this research has used numerical prediction tasks (see Camerer and Johnson 1991; Dawes and Corrigan 1974; Einhorn 1972). Researchers have less frequently used the types of action-based tasks that constitute much (arguably most) of the learning experiences encountered by experts in real situations. Thus, our research indicates that this previous literature may suffer from measurement bias. This bias would arise if naturally occurring expertise resulted from action-based learning. We would expect experts to perform poorly on the typical numerical prediction tasks but well on categorical tasks that better match their learning goals. Therefore, when assessing expert performance, researchers should use tasks and measures that are ecologically valid and are representative of the expert's learning goals, unless they are specifically investigating the transfer of expertise. Comparing categorical and numerical prediction tasks for naturally occurring domains of expertise is a promising area for further research.

Conclusion

In this article, we examined the costs and benefits of learning that result from repeated decision making with outcome feedback (i.e., action-based learning). Our results suggest that learning in naturalistic tasks is not independent of the decisions that lead to that learning. For example, we found that both overall performance and the pattern of performance across stimuli were strongly affected by learning goals. Some goals direct attention toward information that results in learning that transfers across situations, but other goals result in learning that is distorted by the characteristics of the stimuli that were considered the most goal relevant. Contrary to popular wisdom, we found that reliance on this type of experiential learning is likely to be a risky proposition because it can be either accurate and efficient or errorful and biased.

Our results add to the large body of research that has examined the learning of continuous and discrete functions of multiattribute stimuli, and this article is one of a small number of studies that combines these paradigms. The attention-band model is new to the literature and is consistent with our data; however, further research should extend and test the attention-band model relative to traditional models. Another important area for further research is to validate our laboratory results with field studies of action-based learning. Finally, we hope that this research draws attention to the value of better understanding the interplay among goals, attention, and learning in the marketplace.

REFERENCES

Alba, Joseph W. and J. Wesley Hutchinson (1987), "Dimensions of Consumer Expertise," *Journal of Consumer Research*, 13 (4), 411–54.

Ashby, F. Gregory and Ralph E. Gott (1988), "Decision Rules in the Perception and Categorization of Multidimensional Stimuli," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14 (1), 33–53.

——— and W. Todd Maddox (2005), "Human Category Learning," *Annual Review of Psychology*, 56 (1), 149–78.

Brehmer, Berndt and Annica Brehmer (1988), "What Have We Learned About Human Judgment from Thirty Years of Policy Capturing?" in *Human Judgment: The SJT View*, Berndt Brehmer and C.R.B. Joyce, eds. Amsterdam: Elsevier, 75–114.

Camerer, Colin F. (1982), "The Validity and Utility of Expert Judgment," doctoral dissertation, University of Chicago, 1981. Abstract in *Dissertation Abstracts International*, 42, 4171.

——— and Eric J. Johnson (1991), "The Process-Performance Paradox in Expert Judgment: How Can Experts Know So Much and Predict So Badly?" in *Toward a General Theory of Expertise: Prospects and Limits*, K. Anders Ericsson and Jacqui Smith, eds. New York: Cambridge University Press, 195–217.

Cohen, Joel B. and Kunal Basu (1987), "Alternative Models of Categorization: Toward a Contingent Processing Framework," *Journal of Consumer Research*, 13 (4), 455–72.

Dawes, Robyn M. and Bernard Corrigan (1974), "Linear Models in Decision Making," *Psychological Bulletin*, 81 (2), 95–106.

DeLosh, Edward L., Jerome R. Busemeyer, and Mark A. McDaniel (1997), "Extrapolation: The Sine Qua Non for Abstraction in Function Learning," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23 (4), 968–86.

Einhorn, Hillel J. (1972), "Expert Measurement and Mechanical Combination," *Organizational Behavior and Human Decision Processes*, 7 (1), 86–106.

———, Don N. Kleinmuntz, and Benjamin Kleinmuntz (1979), "Linear Regression and Process-Tracing Models of Judgment," *Psychological Review*, 86 (5), 465–85.

Eisenstein, Eric M. (2003), "Action-Based Reasoning: The Cost of Learning and the Benefit of Thinking Less," doctoral dissertation, University of Pennsylvania, 2002. Abstract in *Dissertation Abstracts International*, 63, 4014A.

Goldstone, Robert L. and Mark Steyvers (2001), "The Sensitization and Differentiation of Dimensions During Category Learning," *Journal of Experimental Psychology: General*, 130 (1), 116–39.

Hammond, Kenneth R. (1955), "Probabilistic Functioning and the Clinical Method," *Psychological Review*, 62, 255–62.

Hastie, Reid (2001), "Problems for Judgment and Decision Making," *Annual Review of Psychology*, 52, 653–83.

Hoch, Stephen J. and David A. Schkade (1996), "A Psychological Approach to Decision Support Systems," *Management Science*, 42 (1), 51–64.

Huffman, Cynthia and Michael J. Houston (1993), "Goal-Oriented Experiences and the Development of Knowledge," *Journal of Consumer Research*, 20 (2), 190–207.

Hutchinson, J. Wesley and Joseph W. Alba (1991), "Ignoring Irrelevant Information: Situational Determinants of Consumer Learning," *Journal of Consumer Research*, 18 (3), 325–45.

——— and ——— (1997), "Heuristics and Biases in the 'Eyeballing' of Data: The Effects of Context on Intuitive Correlation Assessment," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23 (3), 591–621.

———, Wagner A. Kamakura, and John G. Lynch (2000), "Unobserved Heterogeneity as an Alternative Explanation for 'Reversal' Effects in Behavioral Research," *Journal of Consumer Research*, 27 (3), 324–44.

Johnson, Eric J. and Jay E. Russo (1984), "Product Familiarity and Learning New Information," *Journal of Consumer Research*, 11 (1), 542–50.

Juslin, Peter, Sari Jones, Henrik Olsson, and Anders Winman (2003), "Cue Abstraction and Exemplar Memory in Categorization," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29 (5), 924–41.

———, Henrik Olsson, and Anna Carin Olsson (2003), "Exemplar Effects in Categorization and Multiple-Cue Judgment," *Journal of Experimental Psychology: General*, 132 (1), 133–56.

Kruschke, John K. (1992), "ALCOVE: An Exemplar-Based Connectionist Model of Category Learning," *Psychological Review*, 99 (1), 22–44.

Levin, Irwin P. (1974), "Averaging Processes and Intuitive Statistical Judgments," *Organizational Behavior and Human Decision Processes*, 12 (1), 83–91.

Logan, Gordon D. (2004), "Cumulative Progress in Formal Theories of Attention," *Annual Review of Psychology*, 55 (1), 207–34.

Maddox, W. Todd, F. Gregory Ashby, and Lawrence R. Gottlob (1998), "Response Time Distributions in Multidimensional Perceptual Categorization," *Perception and Psychophysics*, 60 (4), 620–37.

Medin, Douglas L. and Marguerite M. Schaffer (1978), "Context Theory of Classification Learning," *Psychological Review*, 85 (3), 207–238.

Mellers, Barbara A. (1980), "Configurality in Multiple-Cue Probability Learning," *American Journal of Psychology*, 93 (3), 429–43.

———, Alan Schwartz, and Alan D.J. Cooke (1998), "Judgment and Decision Making," *Annual Review of Psychology*, 49 (1), 447–77.

Meyer, Robert J. (1987), "The Learning of Multiattribute Judgment Policies," *Journal of Consumer Research*, 14 (2), 155–73.

Moreau, C. Page, Arthur B. Markman, and Donald R. Lehmann (2001), "'What Is It?' Categorization Flexibility and Consumers' Responses to Really New Products," *Journal of Consumer Research*, 27 (4), 489–98.

Nosofsky, Robert M. (1984), "Choice, Similarity, and the Context Theory of Classification," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10 (1), 104–114.

———, Thomas J. Palmeri, and Stephen C. McKinley (1994), "Rule-Plus-Exception Model of Classification Learning," *Psychological Review*, 101 (1), 53–79.

Payne, John W., James R. Bettman, and Eric J. Johnson (1993), *The Adaptive Decision-Maker*. Cambridge, UK: Cambridge University Press.

Shafir, Eldar and Robyn A. LeBoeuf (2002), "Rationality," *Annual Review of Psychology*, 53, 491–517.

Sujan, Mita (1985), "Consumer Knowledge: Effects on Evaluation Strategies Mediating Consumer Judgments," *Journal of Consumer Research*, 12 (1), 31–46.

——— and James R. Bettman (1989), "The Effects of Brand Positioning Strategies on Consumers' Brand and Category Perceptions: Some Insights from Schema Research," *Journal of Marketing Research*, 26 (November), 454–67.

Tversky, Amos, Shmuel Sattath, and Paul Slovic (1988), "Contingent Weighting in Judgment and Choice," *Psychological Review*, 95 (3), 371–84.

Van Osselaer, Stijn M.J. and Joseph W. Alba (2000), "Consumer Learning and Brand Equity," *Journal of Consumer Research*, 27 (1), 1–16.

——— and Chris Janiszewski (2001), "Two Ways of Learning Brand Associations," *Journal of Consumer Research*, 28 (2), 202–223.

Weber, Daniel L., David M. Green, and R. Duncan Luce (1977), "Effects of Practice and Distribution of Auditory Signals on Absolute Identification," *Perception and Psychophysics*, 22 (3), 223–31.

West, Patricia M. (1996), "Predicting Preferences: An Examination of Agent Learning," *Journal of Consumer Research*, 23 (1), 68–80.