Action-Based Learning:

Goals and Attention in the Acquisition of Market Knowledge

Eric M. Eisenstein

J. Wesley Hutchinson

Eric M. Eisenstein is an Assistant Professor at the Johnson School, Cornell University and J. Wesley Hutchinson is the Stephen J. Heyman Professor of Marketing, the Wharton School, University of Pennsylvania. We thank Jonathan Baron, Eric Bradlow, Jerry Busemeyer, Teck-Hua Ho, Stephen Hoch, Barbara Mellers, Stijn van Osselaer, Vithala Rao, and three anonymous reviewers for their helpful comments on earlier versions of this paper. Portions of this research were submitted as part of a doctoral thesis at the University of Pennsylvania. The Wharton School funded this research.

Correspondence concerning this article should be addressed to
Eric M. Eisenstein
Johnson School, Cornell University
326 Sage Hall, Ithaca, New York 14850-6201
Phone: 607-255-8627
Fax: 607-254-4590

Email: eric.eisenstein@johnson.cornell.edu

Action-Based Learning:

Goals and Attention in the Acquisition of Market Knowledge

In this paper, we examine the costs and benefits of *action-based learning* (i.e., learning that occurs as a byproduct of making repeated decisions with outcome feedback). We 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, we demonstrate that relying on experiential learning is likely to be a risky proposition because it can be either accurate and efficient or errorful and biased compared to traditional learning.

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 byproduct of repeatedly making decisions about concrete actions and then observing the outcomes. Thus, such learning is *action-based.* 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 byproduct, they form impressions about the effectiveness of their actions. Consumers shop for products, observe prices, make purchases, and assess satisfaction. As a byproduct, they form impressions about the relationship between product attributes and market value. In such everyday environments, most of our experiences afford many dimensions of potential learning and many forms of feedback. Some feedback will be important for immediate goals and some will be important for developing general knowledge; however, there is no guarantee that these two types of feedback will 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 paper, 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 2002; Juslin, Olsson, and Olsson 2003b; 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 allows us to delineate the circumstances under which decision-makers will be most at risk.

Action-Based Learning

The connection between decision-making and learning is that the decision maker often has to make predictions, based on available inputs, about the outcomes that will result from the actions 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 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 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 or not. The discrete action is whether or not to further consider the house (i.e., contact the realtor, visit the property, etc.). This is a *categorical prediction task* (Medin and Schaffer 1978; Tversky et al. 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 the research reported here, 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

---

[1] We have adopted the terminology of Tversky, Sattath, and Slovic (1988) and contrast numerical prediction with categorical prediction. Juslin, Olsson, and Olsson (2003b) use the terms categorization and multiple-cue judgment. We prefer the former because it highlights both the similarity (i.e., prediction) as well as the differences between the two tasks (i.e., continuous vs. discrete responses).

propose a general model in which the market environment creates decision *tasks* that are defined in terms of possible actions. These tasks invoke different decision maker goals, 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 invoke classification-based measures of accuracy (e.g., percent correct). We propose that increasing levels of attention and cognitive effort will be 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 one's knowledge and thereby reduce future error (in ways discussed in more detail later). However, this updating process will not be uniform. It will reflect the definitions of error and the learning goals induced by the decision task. In this sense, the learning is action-based.

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

Learning Categorical vs. Continuous Responses

There has historically been a split between research on the learning of continuous and discrete functions of multiattribute stimuli (DeLosh et al. 1997; Juslin et al. 2003b). 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 vs. exemplar-based models, see Ashby and Maddox 2005; DeLosh et al. 1997; Kruschke 1992; Medin and Schaffer 1978; Nosofsky 1984). In marketing, there have been studies using the categorization paradigm that have investigated consumer learning of brand or category membership based on categorical cues (e.g., Huffman and Houston 1993; Hutchinson and Alba 1991; van Osselaer and Alba 2000). Also, related learning models have been used 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 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 using the multiple-cue probability learning (MCPL) paradigm. MCPL research has shown that a wide variety of functional relationships between cues and outcomes can be learned, given sufficient trials and appropriate feedback, but that some relationships are more difficult to learn than others (see Brehmer and Brehmer 1988 for a review; see also Camerer 1981; Mellers 1980). In particular, linear functions with positive coefficients and no interactions are easiest to learn, and interactions are very difficult to learn under most circumstances. In marketing, the MCPL paradigm has been used by a number of researchers. Meyer (1987) demonstrated that consumers could quickly learn positive attribute-quality relationships in a multiattribute setting. West (1996) used 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 uninformative feedback. Hutchinson and Alba (1997) demonstrated a framing effect in a budget allocation task where past allocations and their sales results served as the stimuli and feedback. Hoch and Schkade (1996) used a managerial credit-forecasting and learning task to investigate the design of decision support systems.

Although there are large literatures on categorization and multiple-cue judgment, they seldom overlap, and the two experimental paradigms are seldom combined in the same study (see DeLosh et al. 1997; Eisenstein 2002; Juslin et al. 2003b for notable exceptions).

Judgment, Choice, and Learning

The distinction we have made between numerical and categorical prediction is similar to the distinction often made between judgment and choice, respectively (e.g., Johnson and Russo 1984; Payne, Bettman, and Johnson 1993; Tversky et al. 1988); however the judgment vs. 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 et al. 1993; Shafir and LeBoeuf 2002). By contrast, research using the prediction paradigm assesses the correspondence between beliefs and the correct answers (e.g., Brehmer and Brehmer 1988; Camerer 1981; Medin and Schaffer 1978). A second major difference between judgment vs. choice and prediction concerns 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 prior to data collection. These preference paradigms stand in contrast to those that focus on learning or the development of conceptual knowledge, which emphasize performance gains that apply to new as well as old informational inputs, provide objective feedback during a training period, and attempt to understand the process through which concepts are added to one's general base of knowledge.

In spite of these substantial differences between the domains, they are clearly similar when regarded as forms of multi-attribute decision making. In fact, Tversky, Sattath, and Slovic (1988; discussed in more detail later) explicitly noted this similarity and showed how their contingent weighting model could account for results from both preference-based and prediction tasks. Although the prediction data they used was collected without feedback, and their model did not attempt to account for learning, the decision process they described for numerical and categorical tasks is easily extended to address our results. We discuss this in greater detail later in the paper.

Attention-Band Model of Action-Based Learning

As discussed earlier, 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 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 given a categorical prediction task have two possible strategies. First, they might 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. Alternatively, their strategy might 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, whether or not 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 they differ qualitatively, the proactive and reactive mechanisms are not mutually exclusive, and both result in selective attention on stimuli near the cutoff.

There is a long tradition of assessing learning 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 (vs. self-reported) measures of beliefs about how stimulus attributes are related to the criterion, similar to part-worths in conjoint analysis. We develop an extension of the cue abstraction model 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 $\mathbf{X}_t = (X_{1t}, X_{2t}, \ldots)$ 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, the learner is assumed to minimize deviations from this goal, rather than deviations from feedback. One very general specification of $G(Y_t)$ is given below.

[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.

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

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 will call the *attention-band*.[3] Within this range, the goal is to learn to predict specific values, so $G(Y_t) = Y_t$. Outside of 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 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 is categorical prediction with a cutoff of Y_c' (and Y_c' may be different from Y_c), then the response to stimulus $\mathbf{X}_t$ is "above," if $R(\mathbf{X}_t) \geq Y_c'$, and "below," if $R(\mathbf{X}_t) < Y_c'$, where $R(\mathbf{X}_t)$ is computed using the final values of $\boldsymbol{\beta}_t$ from the training phase. Analogously, if the test task is numerical prediction, then the response to stimulus $\mathbf{X}_t$ is simply $R(\mathbf{X}_t)$. Because no feedback is given during the test phase of most learning experiments, the model assumes that the weights used

[3] Weber, Green, and Luce (1977) coined this term for a psychophysical model of loudness judgments where 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).

at the end of training are applied at test without further updating.[4]

In some important ways, the attention-band model is similar to the contingent weighting model proposed by Tversky, Sattath, and Slovic (1998). Their model does not consider the role of feedback in learning, but instead focuses on judgments and predictions that are based on multi-attribute inputs. Like 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 their model is driven by the compatibility principle, which holds that when the task is quantitative in nature (e.g., numerical prediction) then quantitative (i.e., continuous) inputs will be more heavily weighted than qualitative (i.e., categorical) inputs and, conversely, when the task is qualitative in nature (e.g., categorical prediction), then qualitative inputs will be more heavily weighted than quantitative inputs. Thus, like the attention-band model, predictions are task-dependent. However, the mechanism they propose is different from the proactive and reactive attention-band mechanisms described earlier. 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 are 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.

H1: *Selective Attention.* Subjects given categorical prediction tasks will allocate increased attention to stimuli near their cutoff compared to numerical prediction subjects.

[4] This assumption was tested with the data from Experiments 2 and 3 for subjects who had matched training and test tasks. No statistically significant differences were found between the estimated coefficients at the end of training and test for any coefficient in either the prediction or cutoff-matched conditions. Details are available from the authors.

This hypothesis follows from the assumptions that the learning goal, $G(Y_t)$, defines error and 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., subjects need only learn Y_L and Y_U for stimuli that are outside of the attention band, but must learn to predict Y_t for stimuli that are inside the attention-band). Subjects given categorical prediction tasks should have relatively narrow attention-bands; subjects in numerical prediction conditions should have very wide attention-bands, effectively allocating attention equally across all stimuli. For example, a manager 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 subjects' ability to discriminate among levels of that attribute; e.g., Goldstone and Steyvers 2001; Medin and Schaffer 1978; Tversky et al. 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 a variety of ways that depend on the stimulus design, which leads hypothesis two.

H2: *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 predicting 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 vs. those in the $600 range. The stimulus designs of our experiments allow us to use H2 to make very specific predictions that differentiate the attention-band model from other models of attention and learning.

H3: *Localized Accuracy*. Compared to numerical prediction trained subjects, categorical prediction trained subjects will exhibit less error for stimuli near their cutoff and more error for stimuli away from their cutoff.

According to the attention-band model, subjects attempt to minimize the discrepancy between $R(\mathbf{X}_t)$ and $G(Y_t)$, so assume $R(\mathbf{X}_t) \approx G(Y_t)$. Because $G(Y_t)$ is constant outside of the attention-band for categorical prediction (i.e., Y_L or Y_U), error (i.e., MAD = $|Y_t - R(\mathbf{X}_t)|$) will be higher outside of the attention band (where $R(\mathbf{X}_t) \approx Y_L$ or $R(\mathbf{X}_t) \approx Y_U$) compared to inside the attention band (where $R(\mathbf{X}_t) \approx Y_t$). For numerical prediction subjects, error should be more uniform across stimuli. Thus, if overall learning is approximately the same for both groups, categorical prediction subjects 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 2002; and Juslin et al. 2003b). We use this paradigm, and its stimulus design, to further refine the general hypotheses. A key aspect of the design was 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, increasing the ecological validity of our experiments. Furthermore, this aspect of the design allowed us to test for the effects of an attention band.

The stimuli used in our experiments were constructed using a multi-attribute model containing 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:

$$
\begin{aligned}
Y \quad = \quad & b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 \\
& b_{13} X_1 X_3 + b_{23} X_2 X_3 + b_{12} X_1 X_2 + b_{123} X_1 X_2 X_3 , \qquad (3)
\end{aligned}
$$

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 *non-overlapping* 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 subjects 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 vs. 48), and when X_3 was +1, the coefficient for X_1 was three times greater than that of X_2 (i.e., 48 vs. 16). This property implies that if only stimuli above or below 500 are considered, then no interactions are present (and X_1 and X_2 have unequal weights), but that 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 below 500 (see below). Also, as mentioned earlier, this design abstracts important characteristics of many real-world markets by using the X_3 to create sub-categories, and by using the interactions to simulate different attribute-price relationships within those market tiers. The specific characteristics of our stimuli also permit us to create more sensitive tests of our hypotheses.

In all experiments, subjects were randomly assigned to a numerical or a categorical prediction task. In a numerical prediction task, subjects predicted price based on the attributes; in a categorical prediction task, subjects classified stimuli as either "above" or "below" a specific price (i.e., a cutoff value on the underlying criterion). Two categorical prediction tasks were used, one in which subjects judged whether the criterion value was above or below 500, and a one in which the cutoff was 325, which we will refer to as the *500-cutoff* task and the *325-cutoff* task, respectively. Equal-length cover stories placed subjects' 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 displayed as either "Above Average" or "Excellent." Experiments consisted of a training phase followed by a test. During training, subjects in all conditions were sequentially shown 72 stimuli about which they had to make 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 subjects made task-appropriate responses. After each trial, subjects 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. 2003a). To facilitate learning, subjects were required to 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, subjects wrote "low" or "high" based on whether their prediction was less than or greater than the actual price.

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

Specific Hypotheses Generated from the Attention-Band Model

The design allows a sensitive test of the *localized learning* hypothesis (H2). We expected that subjects given a categorical prediction task with a low (325) cutoff would define error as incorrect classification, and allocate most of their attention to stimuli below 500 (i.e., the 36 stimuli for which $X_3 = -1$). As a result, they should overweight X_2 and underweight X_1 when compared to subjects given a numerical prediction task. Also, they should underweight the binary attribute because it accounts for no variance among stimuli below 500. Furthermore, the non-overlapping property of our stimuli implies that a categorical prediction task with a cutoff of 500 could be performed based on the binary attribute (X_3) alone. Thus, we expected that subjects given this task would also define error as 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, $R(\mathbf{X}_t)$ was estimated 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. Several specifications of $G(Y_t)$ were used to represent the numerical prediction and cutoff tasks, as prescribed by the attention-band model.[5] The results of these model estimations were consistent with our qualitative reasoning and allowed us to generate, *ex ante,* a robust set of predictions about how learning will be affected by action-based tasks if subjects use strategies that are consistent with the attention-band model. Thus, H2a and H2b are sensitive tests of H2:

H2a: For subjects given a 325-cutoff, there will be an asymmetry in the importance weights for the two continuous attributes, such that $\beta_1 < \beta_2$. Subjects given the other tasks will not exhibit this asymmetry. This will be tested using the difference in weights, $\beta_2 - \beta_1$.

H2b: The weighting of the binary attribute relative to the two continuous attributes will be least

---

[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 from the authors. In these regressions, Y_L was defined to be the mean of all stimuli with criterion values less than L and Y_U was defined 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 (cf., Levin 1974).

for the 325-cutoff task, greatest for the 500-cutoff task, and intermediate for the numerical prediction task. This will be tested 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 noted earlier, the contingent weighting model of preference and choice (Tversky et al. 1988) can be adapted to our task. The compatibility principle predicts that subjects assigned to either cutoff task should give greater weight to the binary attribute than those assigned 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. By contrast, the attention-band model predicts that subjects given the 325-cutoff task will underweight, not overweight, the binary attribute compared to subjects given the prediction task.

Experiment 1

To test H2 - H3, we used the three tasks discussed earlier during the training phase (i.e., 325-cutoff, 500-cutoff, and numerical prediction) and all subjects performed the numerical prediction task in the test phase. H1 will be tested in Experiments 2 and 3.

Method

Subjects. Forty six students enrolled in a consumer behavior course participated for partial course credit and for an opportunity to win a $100 prize. Five subjects were removed from the analysis for failing to follow directions.

Procedure. The general procedure described earlier was used 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). Stimuli were presented to groups of subjects using a computer projector. Therefore, subjects saw all stimuli for the same amount of time and in the same random order. During training, subjects had ten seconds to respond to a presented stimulus by writing their prediction on a paper sheet. 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 ten seconds, the correct numerical answer (i.e., the true price) appeared on the screen for ten seconds, and subjects in the cutoff conditions were also shown the correct task-appropriate category (i.e., above or below the cutoff). In the test phase, all subjects predicted prices and had 12 seconds to respond. Data were again collected on paper sheets.

Results

Localized Learning. OLS regression was used 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 are given in Table 1. H2a predicts that the 325-cutoff subjects will place greater weight on X_2 than on X_1, which was tested using the within-subject difference, $\beta_2 - \beta_1$, as the dependent measure. As predicted, there was a difference for the 325-cutoff condition, $t(13) = -2.21$, $p = .04$, 64% of subjects,[6] but no differences in the other conditions (all t's < .75). As predicted by H2b, $\gamma = \beta_3 / (\beta_1 + \beta_2)$ was least for the 325-cutoff task and greatest for the 500-cutoff task, which was tested by a linear contrast, $F(1, 38) = 9.80$, $MSE = 1.47$, $p = .003$.[7] As discussed earlier, this result violates the compatibility principle of Tversky, et al. (1988).

Insert Tables 1 and 2 and Figure 1 about here

Localized Accuracy. For each participant, we computed the mean absolute deviation (MAD) of the test phase numerical prediction from the true value, the percent correct for the 325-cutoff task (PC_{325}), and the percent correct for the 500-cutoff task (PC_{500}). The categorical predictions implied by the observed numerical responses were used to compute PC_{325} and PC_{500} for each subject. Thus, regardless of training task, each measure of accuracy was computed for each subject.

Hypothesis 3 predicts increased accuracy around the cutoff for subjects in the cutoff

---

[6] Throughout, we report the percentage of subjects 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 1981; Mellers 1980).

conditions. To test this hypothesis, Figure 1 displays each performance measure as a function of the criterion for each condition. To simplify these plots, stimuli were grouped into 7 bins according to criterion values: less than 275, 275 - 370, 370 - 425, 425 - 600, 600 - 675, 675 - 760, and above 760. These boundaries made the number of stimuli in each bin as equal as possible. Figure 1 (MAD panel) reveals that subjects in the cutoff conditions were more accurate (i.e., MAD was lower) near their respective cutoffs than far away and that, by contrast, prediction condition subjects 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 bins. The magnitudes of these values were chosen such that the mean across stimuli was 0 and the spacing between the positive and negative codes was 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. Dummy variables for participants and stimuli were included to statistically control for any general effects of ability or problem difficulty, and stimulus order was used to control for any effect of fatigue. Data were pooled across conditions and MAD was regressed onto *BAND* and the control variables. The results are summarized in Table 2. The significant negative coefficient of *BAND* indicates that cutoff condition subjects had increased accuracy (i.e., lower MAD) near their cutoffs, as predicted by H3.

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

Discussion

The results of Experiment 1 support the attention-band model predictions of localized learning (H2), localized accuracy (H3). The costs and benefits of selective attention were apparent. The enormous advantage in PC_{500} accuracy for 500-cutoff condition subjects (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 subjects and numerical prediction subjects suggests that the former learned that the binary attribute (X_3) perfectly predicted whether price was above or below 500, but the latter did not. The differences in γ (as defined in H2b) reinforce this conclusion. Furthermore, 500-cutoff subjects were more accurate than the numerical prediction subjects in overall MAD and in MAD around the 500 cutoff. By contrast, while the cost of selective attention for the 325-cutoff subjects is clearly 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 the compatibility hypothesis of Tversky, et al. (1988). Overall, the results of Experiment 1 suggest that the costs and benefits of action-based learning will be 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 wished to focus on selective attention (H1). Therefore, we introduced two substantive changes to the design. First, in order to more carefully track the allocation of attention, the experiment was self-paced and response latencies were measured at each stage. H1 predicts that, in cutoff conditions, latencies will be longer for stimuli with prices near the cutoff compared to those with prices far away. Furthermore, self-pacing is more ecologically valid. Second, we manipulated the role adopted by subjects in order to provide an additional test of attentional shifts. In Experiment 1, all subjects 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 subjects adopted the role of a manager and half adopted the role of a consumer. In cutoff conditions, the consumer role was described as an upper limit on spending. This upper limit was expected to shift attention downward toward products that might ultimately be purchased which should increase the overweighting of X_2 relative to X_1. No effect of a role was expected for the numerical prediction condition.

H4: *Role.* The consumer role will shift attention in the categorical prediction conditions toward stimuli with criterion values below the cutoff; therefore, the asymmetry in importance weights for the two continuous attributes (i.e., β_2 - β_1) will become more extreme.

Method

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

Procedure. The general procedure discussed earlier was used with the following specifications. Subjects were randomly assigned to one of the six experimental conditions: 3 (task: 325-cutoff, 500-cutoff vs. numerical prediction) x 2 (role: manager vs. consumer). In order to make the task more relevant for the subject population and to demonstrate the generality of the results obtained in Experiment 1, the product category was changed from houses to digital cameras. The cover story for consumers was screening for affordable products. The managerial cover story was screening for an appropriate sales channel under the assumption that cameras priced over the cutoff would be

[8] The criterion for classification as being a "non-learner" necessarily depended on the experimental cell. For subjects in numerical prediction conditions, achieving an R^2 between the training estimates and the actual prices that was less than 0.20 was used. For the cutoff conditions, failure to achieve a final training accuracy of at least 70% was the criterion.

sold in a different channel than those under 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 the those used in Experiment 1, but were labeled: hours of battery life (X_1), kilo-pixels of resolution (X_2), and optical zoom quality (X_3). Stimuli were presented in six randomized orders.

The training procedure was the same as that used in Experiment 1, except that subjects 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, subjects predicted the values of 80 digital cameras. The first 72 stimuli were the same as those presented during the training phase. These stimuli were followed by an additional eight new stimuli.

Results

Selective Attention. The latency times 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. *Feedback time* is the time spent examining feedback. The average of these latencies are reported in Table 1.[9] Overall, decision time in the numerical prediction condition was longer than 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).

Hypothesis 1 states that subjects assigned to categorical prediction conditions will allocate increased attention to stimuli around the cutoff. This was tested by regressing decision time and feedback time onto *BAND* and the control variables used in Experiment 1. The coefficient for *BAND* was positive and statistically significant for both latency measures, supporting H1 (see Table 2). However, as discussed previously, there are potentially both proactive and reactive mechanisms proposed by the attention-band model that could cause this effect. The reactive mechanism

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

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 only operate for feedback times because subjects do not know their error during decision making. Thus, the significant coefficient of *BAND* for decision times provides evidence for the proactive mechanism. To further differentiate these two psychological mechanisms 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 subjects incorrectly classified the stimulus, and 0 otherwise (i.e., when categorical prediction subjects 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 subjects). Consistent with the reactive mechanism, *C-ERROR* was positive, statistically significant, and relatively large (1.37 seconds, see Table 2). Similarly, subjects in numerical prediction conditions attended to larger errors for a longer amount of time, as 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 for a misclassified stimulus (i.e., 1.37 s). The significant coefficient of *BAND* demonstrates that cutoff subjects allocated .35 seconds of additional feedback time to problems inside the band vs. outside it, even after controlling for the effects of *C-ERROR*, supporting the presence of both proactive and reactive mechanisms.

Localized Learning. As in Experiment 1, OLS regression was used to estimate decision weights for each participant. Mean values for these estimated weights are displayed in Table 1. As in Experiment 1, H2a was tested using the within-subject difference, $\beta_2 - \beta_1$. Replicating the results of Experiment 1, there was a difference in weight in the 325-cutoff condition, $t(16) = -2.39$, $p = .03$, 65% of subjects, but no difference in the numerical prediction condition or in the 500-cutoff condition (all t's < .50). As predicted by H2b, γ was least for the 325-cutoff task and greatest for the 500-cutoff task, which was tested using a linear contrast, $F(1, 46) = 6.09$, $MSE = 2.23$, $p = .02$. As discussed earlier, this violates the compatibility principle (Tversky et al. 1988). Furthermore, as predicted by H4, the role manipulation increased the asymmetry in weights for the two continuous attributes in both cutoff conditions, as shown by contrasting the difference $\beta_2 - \beta_1$ for the cutoff consumer and manager roles (see H5 and Table 1), $F(1, 46) = 5.72$, $p = .02$.

Localized Accuracy. As in Experiment 1, OLS regression was used to estimate the coefficient of *BAND*, which represents the difference in accuracy for stimuli in vs. out of the attention-band for the categorical prediction conditions (see Table 2). The significant negative coefficient of *BAND* indicates that subjects in categorical prediction conditions displayed increased accuracy near their cutoffs, as predicted by H3. As in Experiment 1, the 500-cutoff subjects performed much better than those in other conditions in terms of PC_{500} accuracy around the 500-cutoff. Whereas the 500-cutoff subjects ($PC_{500} = .98$) made essentially no errors in classification around their cutoff, the numerical prediction subjects ($PC_{500} = .86$) were much less accurate for these stimuli, $F(1, 46) = 6.82$, $MSE = .02$, $p = .01$. As in Experiment 1, the subjects 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 H1. Although subjects assigned to categorical prediction tasks spent far less time deciding than did subjects 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, 325-cutoff subjects exhibited the predicted asymmetry in decision weights for the continuous attributes (H2). As predicted, this asymmetry was more extreme for consumer than for manager cutoff roles (H4). Furthermore, both cutoff conditions demonstrated localized accuracy (H3). These results violate the compatibility principle of Tversky, et al. (1988).

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 subjects were equal to numerical prediction subjects 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 subjects were faster than numerical prediction subjects, 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, there is no evidence that self-pacing damaged the learning process in any major way because all results were replicated.

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 subjects in both experiments consistently performed as well as the numerical prediction trained subjects in terms of MAD, and were superior in PC_{500} accuracy. Furthermore, this good performance was also not the result of categorical learning per se (as predicted by the compatibility principle), because it was not observed for the 325-cutoff group. However, in the previous experiments, X_3 was a perfect predictor of the price being above or below 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 reviewed by Ashby and Gott (1988; see also, Nosofsky, Palmeri, and McKinley 1994). To test this conjecture, in Experiment 3 half of the subjects were presented with a stimulus set in which X_3 was a good, but not perfect, predictor of the price being above or below 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 classification could be achieved based solely on X_3. We refer to these as *overlapping* stimuli because both levels of X_3 have some stimuli above and below 500. As in previous experiments, we applied the attention-band model to the overlapping stimuli to generate specific hypotheses (details are available from the authors). The results of these regressions revealed that the near perfect performance of the 500-cutoff condition should be reduced to a level comparable to that of the numerical prediction condition, due to more errors near the cutoff.

H5: *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, all subjects were tested on the numerical prediction task. Although strong evidence for the attention-band model was found, it might be argued that the results were artifacts of the change in task at test or the recoding of the numerical prediction data into categorical predictions. A plausible hypothesis might be that the numerical prediction group did learn that the categorical attribute was a perfect predictor of classification above or below 500, but that without being cued at test, they were unable to recall that helpful piece of knowledge when making their predictions. Using the 500-cutoff task at test 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 was stored in memory but not retrieved in earlier 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 later undone by providing or denying appropriate retrieval cues.

H6: *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 subjects' performance.

Method

Subjects. One hundred fifty four students participated for partial course credit. None had previously participated in similar experiments. A total of twelve subjects were removed from the analysis for failing to follow directions or failing to learn during the training phase, using the same criteria as in Experiment 2.

Procedure. The stimuli and procedures of Experiment 2 were used with the following changes. Subjects were randomly assigned to one of the eight experimental conditions: 2 (training task: 500-cutoff or numerical prediction) x 2 (test task: 500-cutoff or numerical prediction) x 2 (overlap: non-overlapping or overlapping stimuli). The non-overlapping stimuli and digital camera cover stories were identical to those used in Experiment 2 with the exception of cover story changes that were necessary to allow for the cutoff test task. The overlapping stimuli were described earlier. 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 used in Experiment 2, with the exception of the instructions for the 500-cutoff test task.

Results

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

As in Experiment 2, selective attention (H1) was tested 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, supporting H1 (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 discussed earlier. Similarly, subjects in numerical prediction conditions attended to larger errors for a longer amount of time, as 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 for a misclassified stimulus (i.e., 2.02 s). Finally, the significant positive coefficient of *BAND* indicates that subjects in categorical prediction conditions allocated .62 seconds of additional feedback time to problems inside the band vs. outside it, even after controlling for the effects of *C-ERROR*, supporting the presence of both proactive and reactive mechanisms.

Localized Learning. One half of the subjects were tested on the numerical prediction task. As in previous experiments, OLS regression was used 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 are given in Table 1. H2a is not applicable, as there was no 325-cutoff condition. H2b predicts that γ will be greater for 500-cutoff subjects than for prediction subjects. ANOVA revealed a training task x overlap interaction $F(1, 66) = 3.93$, $MSE = .95$, $p = .05$, which is due to the fact that the non-overlap condition strongly showed this difference in γ, but the overlap condition was only directionally consistent.

Localized Accuracy. As in previous experiments, regression was used to estimate the coefficient of *BAND*, which represents the difference in accuracy for stimuli in vs. out of the attention-band for the cutoff condition (see Table 2). The significant negative coefficient of *BAND* indicates that subjects assigned the categorical prediction task displayed increased accuracy near their cutoffs, as predicted by H3. This result held for both the overlapping and non-overlapping

[10] As in Experiment 2, extreme latency times for each subject were removed from the analysis (see footnote 11). Total deleted data was 2.1% of observations, and the average was 1.5 problems deleted per subject.

conditions. Replicating the results of previous experiments, subjects in the 500-cutoff non-overlapping condition made essentially no errors in classification around their cutoff (PC_{500} = .98), but subjects in the numerical prediction condition were much less accurate (PC_{500} = .76), $F(1, 135)$ = 56.3, p < .0001. When the simple verbal rule was removed eliminated by overlapping stimuli, the 500-cutoff trained condition remained superior by about 6 percentage points, $F(1, 135) = 4.22$, p = .04, providing further support for H3.

Simple Rule and Encoding. PC_{500} was computed for all subjects. H5 predicts that the overlapping stimuli will degrade the PC_{500} performance of the 500-cutoff condition more than the prediction condition. H6 predicts that explicitly cuing the 500 boundary by testing on the 500-cutoff task will not improve numerical prediction trained subjects' performance. An ANOVA 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 x overlap interaction, $F(1, 135) = 5.21$, p = .02. As predicted by H5, 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 in any case was not differentially beneficial to the numerical prediction condition (amounting to only a 1% increase in percent correct). This result indicates that unattended information is lost at the time of encoding, and supports H7.

Discussion

Overall, these results replicate and extend the results of the previous experiments. As before, selective attention (H1, both proactive and reactive mechanisms), localized learning (H2), and localized accuracy (H3), were supported. The removal of the simple verbal rule by using overlapping stimuli had the expected effect of degrading the performance of cutoff-trained subjects more than numerical prediction trained subjects, supporting H5. Nonetheless, the cutoff-trained subjects 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 predicted by H6, numerical prediction trained subjects 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 later problems with retrieval.

General Discussion

In the research reported here, 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, as it arises whenever a cutoff-based screening process is followed by evaluation of the considered options. For example, the environment used in these experiments is analogous to consideration set formation, screening new product ideas for further development, and instructing agents about buying and selling prices. An important distinction was made between learning to estimate the criterion (numerical prediction) and learning to correctly classify a stimulus 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 quite 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 was paid to stimuli and to feedback and 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 as for people trained on numerical prediction, even though much less time was invested 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 more effortful, this strategy has the advantage of preparing the learner for new decision tasks that might be encountered in the future. Thus, our results show that there are advantages and disadvantages to action-based learning, because accuracy will depend on whether or not 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 highlights 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. Such situations frequently arise 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 based on 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, a salesperson successfully "up-sells" 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 then 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. More generally, 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 thirty 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, high levels of attention are assumed to be 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 directed by error feedback.

Expertise

In the large literature on expertise, researchers have measured expert performance in a variety of ways, and they have generally found it to be surprisingly low. Much of this research has used a numerical prediction task (see Camerer and Johnson 1991; Dawes and Corrigan 1974; Einhorn 1972). Researchers have less frequently used the types of action-based tasks that comprise much (arguably most) of the learning experiences encountered by experts in real situations. Thus,

[11] We have 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 reported here, except that the original GCM 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 from the authors.

our research suggests 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 future research.

Conclusion

In this paper, we examined the costs and benefits of learning that results 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 seen as most goal-relevant. Contrary to popular wisdom, we found that relying 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 paper is one of a small number of studies that has combined these paradigms. The attention-band model is new to the literature and was consistent with our data; however, it remains for future research to extend and test it relative to traditional models. Another important area for future 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 between goals, attention, and learning in the marketplace.

TABLE 1

RESULTS FOR EXPERIMENTS 1 – 3

		Experiment 1[a]			Experiment 2						Experiment 3			
Measure or parameter	Normative value	325-Cutoff Mgr	500-Cutoff Mgr	Num Pred Mgr	325-Cutoff Mgr	500-Cutoff Mgr	Num Pred Mgr	325-Cutoff Con	500-Cutoff Con	Num Pred Con	500-Cutoff Mgr	Num Pred Mgr	Overlap 500-Cutoff Mgr	Overlap Num Pred Mgr
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
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] $000's for Experiment 1 and dollars for Experiments 2 and 3

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

TABLE 2

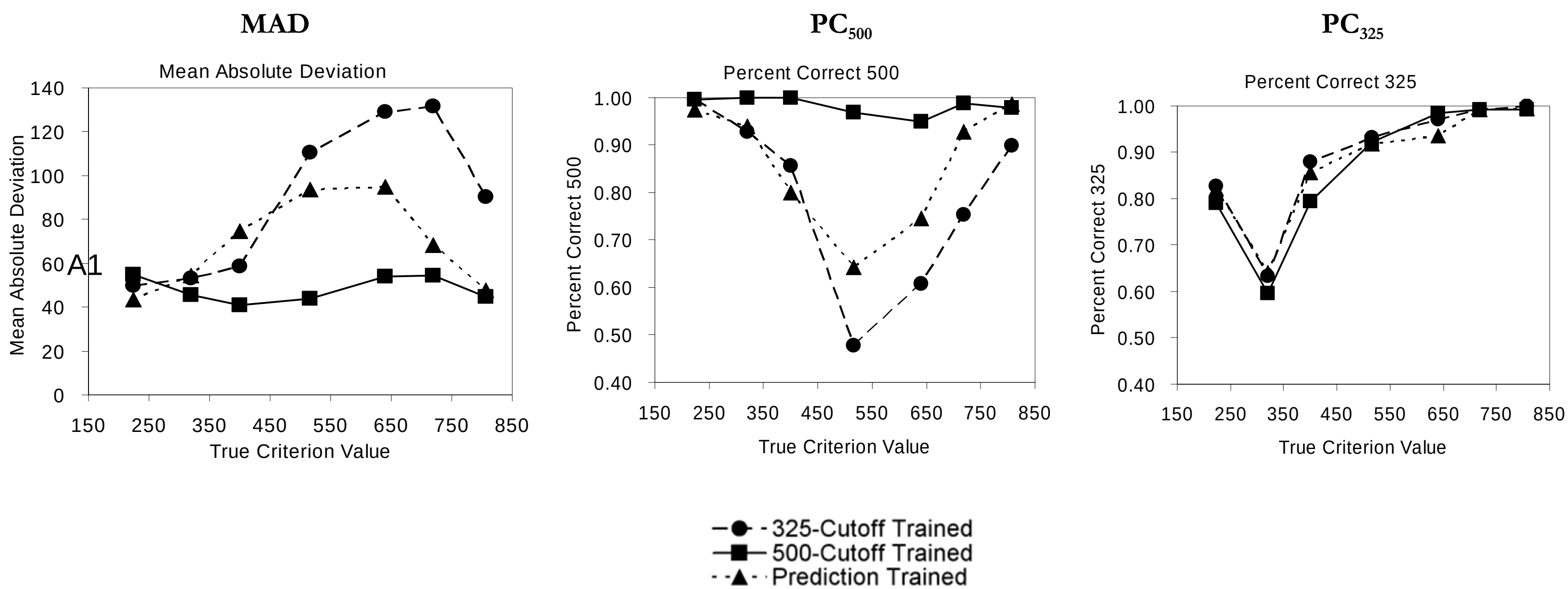
LOCAL ACCURACY AND INCREASED ATTENTIONAL ALLOCATION AROUND THE CUTOFF

Experiment	Measure	BAND	C-ERROR	P-ERROR
1	MAD	-35.47* (5.51)		
2	Feedback Time (s) – No Error	0.52* (.10)		
	Feedback Time (s) – Error	0.35* (.10)	1.37* (.17)	0.0060* (.0009)
	Decision Time (s)	0.62* (.14)		
	MAD	-14.52* (2.64)		
3	Feedback Time (s) – No Error	0.51* (.10)		
	Feedback Time (s) – Error	0.31* (.09)	2.02* (.12)	0.0058* (.0005)
	Decision Time (s)	0.61* (.20)		
	MAD	-12.28* (-3.25)		

NOTE. * indicates $p < .001$. The standard error of each coefficient appears in parentheses. Explanation of items controlled for: Problem ID, a dummy variable for each problem; subject, a dummy variable for each subject; fatigue, a linear term representing the ordinal position in which the problem was presented during training.

FIGURE 1

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



NOTE. MAD estimates are the mean of the median within-participant accuracy within each bin and percent correct measures are the mean of the mean within-participant accuracy within each bin.

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.

Ashby, F. Gregory 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. North-Holland: Elsevier Science Publishers, B.V.

Camerer, Colin F. (1981), "The validity and utility of expert judgment," Ph.D. diss., University of Chicago, Abstract in *Dissertation Abstracts International* 42 (1982): 4171.

Camerer, Colin F. 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, NY, USA: Cambridge University Press.

Cohen, J. B. and K. 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 & Human Decision Processes*, 7 (1), 86-106.

Einhorn, Hillel J., Don N. Kleinmuntz, and Benjamin Kleinmuntz (1979), "Linear regression and process-tracing models of judgment," *Psychological Review*, 86 (5), 465-85.

Eisenstein, Eric M. (2002), "Action-based reasoning: The cost of learning and the benefit of thinking less," Ph.D. diss., University of Pennsylvania, Abstract in *Dissertation Abstracts International* 63 (2003): 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, R. (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 (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.

---- (1991), "Ignoring Irrelevant Information - Situational Determinants of Consumer Learning," *Journal of Consumer Research*, 18 (3), 325-45.

Hutchinson, J. Wesley, 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, E. J. and J. 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 (2003a), "Cue abstraction and exemplar memory in categorization," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29 (5), 924-41.

Juslin, Peter, Henrik Olsson, and Anna Carin Olsson (2003b), "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), 207-38.

Mellers, Barbara A. (1980), "Configurality in multiple-cue probability learning," *American Journal of Psychology*, 93 (3), 429-43.

Mellers, Barbara A., 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.

Nosofsky, Robert M. (1984), "Choice, similarity, and the context theory of classification," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10 (1), 104-14.

Nosofsky, Robert M., 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: Cambridge University Press.

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

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

Sujan, Mita 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 (4), 454-67.

Tversky, Amos, Shmuel Sattath, and Paul Slovic (1988), "Contingent weighting in judgment and choice," *Psychological Review*, 95 (3), 371-84.

van Osselaer, S. J. and J. W. Alba (2000), "Consumer learning and brand equity," *Journal of Consumer Research*, 27 (1), 1-16.

van Osselaer, Stijn M. J. and Chris Janiszewski (2001), "Two ways of learning brand associations," *Journal of Consumer Research*, 28 (2), 202-23.

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.