Costs and Benefits in Perceptual Categorization

W. Todd Maddox and Corey J. Bohil
University of Texas, Austin

2000, *Memory & Cognition, 28,* 597-615.

Observers categorized perceptual stimuli when the category costs and benefits were manipulated across conditions, and costs were either zero or non-zero. The cost-benefit structures were selected so that performance across conditions was equivalent with respect to the optimal classifier. Each observer completed several blocks of trials in each of the experimental conditions, and a series of nested models were applied to the individual observer data from all conditions. In general, performance became more nearly optimal as observers gained experience with the cost-benefit structures, but performance reached asymptote at a sub-optimal level. Observers behaved differently in the zero and non-zero cost conditions, performing consistently worse when costs were non-zero. A test of the hypothesis that observers weight costs more heavily than benefits was inconclusive. Some aspects of the data supported this differential weighting hypothesis, but others did not. Implications for current theories of cost-benefit learning are discussed.

Everyday we make important decisions based on uncertain information. For example, we might decide to "bring" or "not bring" an umbrella to work based solely on uncertain predictors of rain, like the degree of overcast. This is a categorization problem because there are many degrees of overcast that one might observe, but only two potential decisions. There might be no or a few clouds, and we might decide "not to bring" an umbrella to work, or there might be many or a sky full of clouds, and we might decide "to bring" an umbrella to work. Often the costs and benefits associated with each categorization decision are different. If we bring an umbrella and it rains we benefit by staying dry; the benefit might be greater if we are wearing a suit than if we are wearing a t-shirt and jeans. If we fail to bring an umbrella and it rains we get wet. This cost might be greater if we are wearing expensive clothes. Costs and benefits can strongly impact the decisions we make. For example, we might decide to "bring" an umbrella if wearing expensive clothes, or "not bring" an umbrella if wearing a t-shirt and jeans, even though the degree of overcast is nearly identical in both cases.

The perceptual and cognitive processes involved in solving categorization problems of this sort have been studied extensively (Ashby, 1992; Busemeyer & Myung, 1992; Green & Swets, 1966; Macmillan & Creelman, 1991; Maddox, 1995; Maddox & Ashby, 1993; Maddox & Bohil, 1998a, 1998b; Nosofsky, 1986; Stevenson, Busemeyer, & Naylor, 1991; vonWinterfeld & Edwards, 1982). One popular approach has been to compare human performance with that of the optimal classifier—a hypothetical device whose categorization decisions maximize long-run reward. When observed performance is sub-optimal, the goal has been to identify the locus of the performance sub-optimality.

Despite anecdotal evidence that suggests differential costs and benefits have a strong impact on everyday categorization decisions, few categorization studies have manipulated systematically costs and benefits. The few studies that have examined costs and benefits have focused only on unequal benefits while holding costs fixed at zero (e.g., Bohil & Maddox, 1999; Maddox & Bohil, 1998b; however see Busemeyer & Myung, 1992). The effects of non-zero costs have been examined in the decision-making literature, and a common finding is that costs are weighted more heavily than benefits (e.g., Kahneman & Tversky, 1979; see also Higgins, 1987). Nearly all of this work used word problems in which the relevant information was presented explicitly (Kahneman & Tversky, 1973; Tversky & Kahneman, 1974, 1980), whereas a large body of categorization research suggests that categories are better learned implicitly through trial-by-trial experience with category members, and trial-by-trial feedback (e.g., Estes,

---

This research was supported in part by a National Science Foundation Grant SBR-9796206 and NIH Grant R01 MH59196. We thank Greg Ashby and Jerome Busemeyer for many interesting discussions that influenced this work. Special thanks go to Jerome Busemeyer, Thomas Ward and one anonymous reviewer who provided comments on an earlier version of the manuscript. We would also like to thank Candace Blair for help with the data collection. Correspondence concerning this article should be addressed to W. Todd Maddox, Department of Psychology, Mezes Hall 330, University of Texas, Austin, Texas, 78712. E-mail: *maddox@psy.utexas.edu*.

Campbell, Hatsopolous, & Hurwitz, 1989; Holyoak & Spellman, 1993; Koehler, 1996; Maddox & Ashby, 1993). We propose a formal model that instantiates this differential weighting of costs over benefits. We test this Differential Weighting Hypothesis (DWH) by applying the model to the categorization data collected in the current study.

The goals of the current study were many. First and foremost, we attempted to bridge the gap between categorization and decision-making research by manipulating systematically both costs and benefits within the framework of a single categorization task. Although categorization is our primary interest, we tried to draw parallels between these two important areas of research. In particular, when appropriate we tried to relate the different terminology used in the two areas. Second, we examined performance in zero-cost and nonzero-cost conditions, and compared performance with that of the optimal classifier. To facilitate a comparison with the optimal classifier, the costs and benefits were constrained in such a way that the performance of the optimal classifier was identical across conditions. Third, we studied the time-course of cost-benefit learning by having observers complete several blocks of trials in each experimental condition, and by analyzing the data from each block of trials separately. The details of these analyses are reported in the Appendix, and should provide a rich database for testing current and future models of cost-benefit learning in categorization. Finally, we provided an initial test of the DWH as applied to categorization.

The next (second) section describes briefly the optimal classifier, and the theoretical framework (decision bound theory) within which the current study operates. The third and fourth sections review briefly the relevant categorization and decision-making literature. The fifth section outlines the experiment and the methods, and the sixth section is devoted to the results and theoretical analyses. Finally, we conclude with some general comments.

Decision Bound Theory and the Optimal Classifier

Consider the situation facing a medical doctor who must classify a patient into one of two disease categories, A or B. Suppose the patient is given medical test X, which is diagnostic of the two diseases. In addition, suppose that the outcomes of test X for diseases A and B are normally distributed with means μ_A and μ_B, and variances σ^2_A and σ^2_B, as depicted in Figure 1a. The optimal classifier records perfectly the test result, denoted x. [In the decision-making literature, the optimal classifier is derived from the "expected utility rule" (e.g., Stevenson, et al, 1991; Yates, 1990). Once we outline the details of the optimal classifier, we will detail its relation to the expected utility rule and other constructs from the decision-making literature.] In other words, given a fixed physical input, the optimal classifier will show no variability in the perceptual representation[1]. The optimal classifier has perfect knowledge of the distribution of test results for each disease category (i.e, the form and parameters of the distribution). This information is used to construct the *optimal decision function*, which is computed from the likelihood ratio of the two category distributions,

$$l_o(x) = f(x|B) / f(x|A) \qquad (1)$$

where f(x|i) denotes the likelihood of test result x given disease category i. If test result x is more likely to result from disease B than disease A, then the likelihood ratio will be greater than one. Conversely, if the test result is more likely to result from disease A than disease B, then the likelihood ratio will be less than one.

The optimal classifier has perfect knowledge of the costs associated with incorrect diagnoses, and the benefits associated with correct diagnoses. This information is used to construct the *optimal decision criterion*

$$\beta_o = (V_{aA}-V_{bA})/(V_{bB}-V_{aB}) \qquad (2)$$

where V_{aA} and V_{bB} denote the benefits associated with correct diagnoses, and V_{bA} and V_{aB} denote the costs associated with incorrect diagnoses (Letters in lower-case denote responses, and letters in upper-case denote categories). The costs and benefits might vary depending on the severity of the disease. For example, suppose that disease A is life threatening, whereas disease B is not. If the doctor correctly diagnoses the patient as suffering from disease "A", the benefit could be large, such as an early diagnosis that saves the patient's life. A correct disease "B" diagnosis, on the other hand, might yield a smaller benefit. Incorrectly diagnosing the patient as suffering from disease "B" (non-life threatening) instead of disease "A" (life

[1] Of course, there may be random variation in the physical input to the perceptual system from a fixed stimulus. For example, there might be error in the measuring device.

threatening) could have a large cost, such as a loss of life. Although potentially "nerve racking", the cost of an incorrect disease "A" diagnosis would likely be less severe.

The optimal classifier (e.g., Green & Swets, 1966) uses $l_o(x)$ and β_o to construct the *optimal decision rule (derived from expected utility theory)*:

$$\text{If } l_o(x) > \beta_o, \text{ then respond "B", otherwise respond "A".} \quad (3)$$

When the costs and benefits lead to no bias, $\beta_o = 1$. In other words, the diagnosis would be driven completely by the likelihood ratio, as depicted by the $\beta_o = 1$ decision criterion in Figure 1a. If the likelihood of disease A exceeds that of disease B, then the patient will be diagnosed with disease "A". If the likelihood of disease B exceeds that of disease A, then the patient will be diagnosed with disease "B". When $V_{aA} - V_{bA} > V_{bB} - V_{aB}$, $\beta_o > 1$. For example, suppose $\beta_o = 3$, as depicted in Figure 1a. Under these conditions, the likelihood of disease B would have to be at least 3 times larger than the likelihood of disease A for the patient to be diagnosed with disease "B". This is a possible scenario if disease A is life-threatening, and disease B is not. It would be much worse to incorrectly diagnose the patient as suffering from a non-life-threatening disease (when their life was actually in danger), than it would be to incorrectly diagnose the patient as suffering from a life-threatening disease (when their life was not actually in danger). The partition between the "A" and "B" response regions [where $l_o(x) = \beta_o$] is called the *optimal decision bound*.

The optimal decision bound is constructed from the "objective" or "true" category information. For example, the optimal classifier has knowledge of the "objective" costs and benefits. A large body of decision-making research suggests that people do not use the objective costs and benefits, but rather base their decisions on subjective costs and benefits that are directly related to the objective values (e.g., Stevenson, et al, 1991; Yates, 1990). To distinguish the subjective values from the objective values, decision theorists use the term utility, to denote the subjective worth of an objective value. Thus, within the framework of decision theory, each of our V_{iJ} terms (where i = a or b, and J = A or B) should be converted into a subjective utility denoted $u(V_{iJ})$, where u describes the functional relationship between the subjective and objective values. When the costs and benefits are quantitative, as is the case with money, it is reasonable to assume that more is always preferred to less. In other words, the utility of a given outcome is monotone increasing with the objective value of the outcome. This is a fairly weak constraint on the functional relationship between utilities and objective values. As a result many functions are possible. However, decision theorists have focused on only a few. In particular, concave, convex, and linear relationships have been investigated. Although a detailed examination of these functional relations and their implications for decision-making and categorization problems is beyond the scope of this article, the linear function is most directly relevant to the current study. Because the optimal decision criterion, β_o, is derived from the ratio of intervals, it is invariant under any linear transformation. In other words, as long as the utility function, u, is linear, then the optimal decision criterion is equivalent to the expected utility criterion. Thus, within the framework of decision theory, the focus of the current article is on linear utility functions. This is a strong assumption, but one that should hold (approximately) for the small range of monetary values in the present studies. For ease of exposition we will generally use the term optimal decision criterion, but when relevant and informative we will relate the discussion back to decision theory.

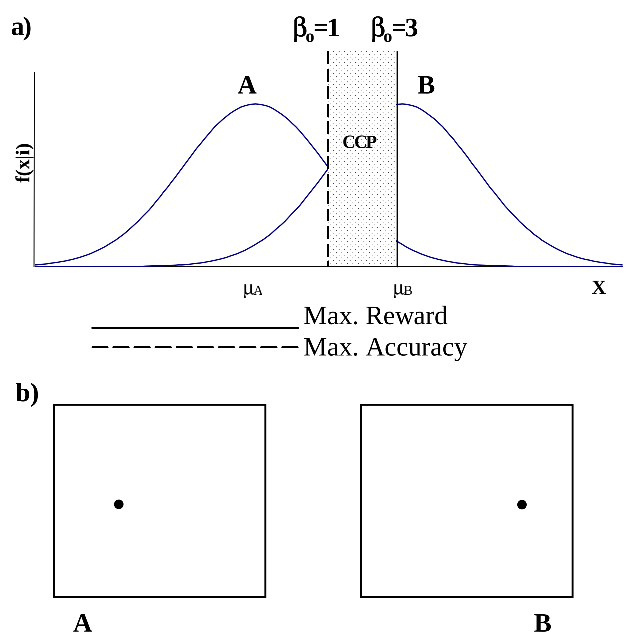

Figure 1. (a) Decision criteria for the classifier that maximizes long-run reward ($\beta_o = 3$) and long-run accuracy ($\beta_o = 1$) in the experimental conditions of the current experiment. The shaded area denotes the region for which conservative cutoff placement results; (b) Representative stimuli from each category.

The optimal classifier decision rule (Equation 3) has been rejected as a model of human performance, but in many cases, performance approaches that of the optimal classifier as the observer gains experience with the task. Ashby and colleagues argued that the observer attempts to respond in accordance with the optimal classifier, but fails because of various sub-optimalities in perceptual and cognitive processing (Ashby, 1992; Ashby & Lee, 1991; Ashby & Maddox, 1993, 1994; Ashby & Townsend, 1986; Maddox & Ashby, 1993; Thomas, 1995). Ashby (1992; see also Maddox & Ashby, 1993) proposed a series of decision bound models that can be used to test specific hypotheses about the locus of performance sub-optimalities. Two sub-optimalities that are inherent in humans (and other organisms), are perceptual and criterial noise. Perceptual noise exists because there is trial-by-trial variability in the perceptual information associated with each stimulus. Thus, the observer's percept of Stimulus i, on any trial, is given by $x_{pi} = x_i + e_p$, where x_i is the observer's mean percept, and e_p is a random variable that represents the effects of perceptual noise. At the cognitive level, there is trial-by-trial variability in the observer's memory for the decision criterion (termed *criterial noise*)[2]. Because of criterial noise, the decision criterion used on any trial is given by $\beta_c = \beta + e_c$, where β is the observer's average decision criterion, and e_c is a random variable that represents the effects of criterial noise (assumed to be univariate normally distributed).

Because perceptual and criterial noise exist, the human observer can not attain the level of performance reached by the optimal classifier (i.e., can not maximize long-run reward). Even so, decision bound theory assumes that the observer attempts to use the same *strategy* as the optimal classifier, but with less success due to the effects of perceptual and criterial noise (and other possible sub-optimalities). Hence, the simplest decision bound model is the *optimal decision bound model*. The optimal decision bound model is identical to the optimal classifier (Equation 3) except that perceptual and criterial noise are incorporated into the decision rule. Specifically,

$$\text{If } l_o(x_{pi}) > \beta_o + e_c \text{ then respond "B", otherwise respond "A".} \tag{4}$$

Besides perceptual and criterial noise, other sub-optimalities might exist. For example, sub-optimalities might exist in category distribution knowledge. Continuing with our medical diagnosis example, the doctor may not know the distribution parameters that describe the relation between medical test X and the two diseases (i.e., μ_A, μ_B, σ^2_A, and σ^2_B). Thus, instead of using the optimal decision function l_o, the observer might use a sub-optimal decision function. Sub-optimalities might also exist in knowledge of the costs and benefits, or in the way this information is combined. For example, the utility function might not be linear, or the doctor might underestimate the true costs associated with misdiagnoses. Thus, instead of using the optimal decision criterion, the observer might use a sub-optimal decision criterion.

Recently Maddox and Bohil (1998b; Bohil & Maddox, 1999) examined performance in simple analogs of the medical decision making problem outlined above. The focus was on an examination of categorization performance when the benefit associated with a correct response for one category was larger than the benefit associated with a correct response for the other category. In all cases, the cost associated with an incorrect response was zero. Simple two-dimensional stimuli were utilized, and observers completed a large number of sessions to ensure performance had reached asymptote. The analyses were conducted at the level of the individual observer, and centered on the application of three decision bound models, each of which made a different assumption regarding sensitivity to the benefit manipulation. The *optimal decision bound model* (OPT; Equation 4) assumed that the observer had accurate knowledge of the category benefits, and thus used the optimal decision criterion, β_o (albeit in the presence of criterial noise) in an attempt to maximize long-run reward. The *payoff neglect model* (PON) assumed that the observer ignored the true category benefits, and instead behaved as if the benefits were equal (i.e., assumed $\beta = 1$). An interesting facet of cost-benefit manipulations is that they require the observer to trade off accuracy maximization (which is always driven by the equal likelihood criterion), and reward maximization. Thus, another way to think of the PON model is as an accuracy-maximization model.

[2] Throughout this article, it is assumed that the perceptual variance, $\sigma_{pi} = \sigma_p$. In many cases these assumptions are incorrect (e.g., Ashby & Lee, 1991, 1993; Maddox & Ashby, 1996, 1998; Maddox & Bogdanov, 1999), but with high contrast, response-terminated displays, and fairly simple stimuli, as in the current study, this is often a reasonable assumption. With stimulus dimensions that yield complex perceptual representations, such as hue, saturation, and brightness, or with brief stimulus displays, these assumptions would be unsatisfactory.

Within the framework of decision theory, the PON is equivalent to a situation in which the utility function is equal to one for correct responses, and is equal to zero for incorrect responses. We elaborate on this issue in the General Discussion. [Throughout this article we will use the term “payoff” to denote the global effects of a cost-benefit manipulation. For example, if the cost-benefit structure is such that the observer is biased toward response “A”, then we refer to category A as the High Payoff category. Thus, if the observer is ignoring the cost-benefit structure, then they are defined as showing payoff neglect.] The *sub-optimal payoff model* (S-O Payoff) assumed that the observer was sensitive to the difference between category benefits, but used a sub-optimal decision criterion. [There are several possible interpretations of this model within the framework of decision theory. One possibility is that the utility function is piecewise linear but the slope and intercept differs for costs and benefits. Another possibility is that the utility function is nonlinear.] In this model, the value of β was estimated from the data. Because β was a free parameter in the S-O Payoff model, but was fixed a priori in the OPT and PON models, the OPT and PON models are “nested” under the S-O Payoff model. The models are “nested” because the OPT and PON models can be derived from the S-O Payoff model by setting the β parameter to a fixed value. Likelihood ratio tests (Ashby, 1992; Wickens, 1982) were used to identify the most parsimonious model for each observer. [Single observer analyses were performed because Maddox (1999) showed that averaging can alter the structure of the data in such a way that the correct model of individual performance can provide a poor account of averaged performance.] The most parsimonious model is defined as the model with the fewest number of free parameters that is not "significantly" improved upon by a more general model. The same procedures were used in the current study.

To summarize, in Maddox and Bohil (1998b) all of the observers showed some sensitivity to the category benefit difference (i.e., the PON model was rejected for every observer), but few used the optimal decision criterion. An examination of the β estimates from the S-O Payoff model suggested that observers tended to use a decision criterion that was more conservative than the optimal decision criterion. For example, when the benefits were such that $\beta_o = 3$, observers tended to use a β between 1 and 3. This is termed *conservative cutoff placement*, and is depicted in Figure 1a by the shaded area. This is a common finding in the literature (e.g., Busemeyer & Myung, 1992; Green & Swets, 1966; Healy & Kubovy, 1981).

Cost and Benefit Manipulations in Categorization and Decision-Making

Categorization studies that manipulate the category benefits while setting the category costs to zero indicate that observers tend to show conservative cutoff placement. Unfortunately, with one exception (Busemeyer & Myung, 1992), no studies have examined the effects of non-zero costs on human categorization performance. This is an important issue because a large body of decision-making research suggests that observers are not only sensitive to cost-benefit manipulations, but also that observers tend to place more weight on costs than on benefits (e.g., Kahneman & Tversky, 1979; see also Higgins, 1987). As stated earlier, most of these studies used word problems in which the relevant information was presented explicitly. One goal of the present study will be to formalize this DWH and to test it in a traditional categorization task where the relevant information is learned implicitly through experience with the category members. Several instantiations of the DWH are possible, but we thought that a reasonable approach was to generalize the equation for the optimal decision criterion (Equation 2). Specifically, we assumed that a decision weight, w_A, was assigned to the benefit of a correct “A” response, and $1-w_A$ was assigned to the cost of an incorrect “B” response to an exemplar from Category A. Analogously, a decision weight, w_B, was assigned to the benefit of a correct “B” response, and $1-w_B$ was assigned to the cost of an incorrect “A” response to an exemplar from Category B. In other words,

$$\beta_{DWH} = \frac{w_A V_{aA} - (1 - w_A) V_{bA}}{w_B V_{bB} - (1 - w_B) V_{aB}} \qquad (5)$$

where $0 \le w_A \le 1$ and $0 \le w_B \le 1$. Of course, it is reasonable to suppose that $w_A = w_B$. Several comments are in order. First, when $w_A < .5$ and $w_B < .5$ the costs are being weighted more heavily than the benefits, and the DWH is supported. Second, when $w_A = .5$ and $w_B = .5$, $\beta = \beta_o$, and so the optimal decision criterion is a special case of the differential weighting decision criterion (c.f., Equation 2 with Equation 5). Third, within the framework of decision-making theories, the differential weighting rule is a generalization of expected value theory, but it is still a special case of expected utility theory.

As an initial test of the DWH we compared performance across a number of cost-benefit conditions for which the DWH predicted qualitatively different patterns of responding. Specifically, we constructed "equivalent" cost-benefit conditions for which the DWH predicted conservative cutoff placement in some conditions, and extreme cutoff placement in other conditions. [Extreme cutoff placement results when the decision criterion is larger than the optimal value.] By "equivalent" we mean conditions for which the predictions of the optimal classifier (i.e., the optimal decision criterion and the optimal long-run reward) are identical. The details will be discussed shortly.

EXPERIMENT

The goals of the present research were three-fold. First, we hoped to bridge the gap between the fields of categorization and decision-making by examining several categorization problems in which the cost-benefit structure was manipulated. Second, we compared and contrasted decision criterion learning across situations for which the cost of an incorrect response was either zero or non-zero. These learning data should provide a rich database for current and future model testing. Finally, we provided an initial test of the DWH in categorization. To achieve these goals we had observers complete several categorization problems, each with a unique cost-benefit structure. Although the costs and benefits differed across conditions, the category structures (i.e., the category means and variances), optimal decision criterion (β_o = 3), and maximum (optimal) reward were held fixed across conditions. The costs and benefits, and DWH predictions (whether conservative or extreme cutoff placement) for different combinations of w_A and w_B for each of the five experimental conditions are outlined in Figure 2.

Notice that the optimal decision criterion, β_o, is equal to 3 in each condition. This can be verified by applying Equation 2 to each of the five cost-benefit structures, or by examining the DWH predictions when $w_A = w_B = .5$. Notice also that each cost-benefit condition has an associated multiplicative scalar, α. The value of α for each condition was selected in such a way that the maximum (optimal) reward was (approximately) equal in each condition. We thought it important to equate the maximum reward to ensure that observers were equally motivated in each experimental condition. Figure 2a displays the costs and benefits and DWH predictions for the "Zero Cost" condition. This served as a control condition against which the four "Non-zero Cost" conditions (Figure 2b – 2e) were compared. In this condition, the DWH predicts conservative cutoff placement anytime $w_A < w_B$. Notice that the costs and benefits for the Figure 2b – 2e conditions are related systematically. For example, in Figure 2b the cost of an incorrect response was equal across both categories (-2α) but the benefit of a correct category "A" response (4α) was higher than the benefit of a correct category "B" response (0α). The Figure 2c cost and benefit entries are very similar to those from Figure 2b except that the numerical values in each column (excluding the sign) were switched. This leads to an important change in the predictions from the DWH. Specifically, when costs are weighted more heavily than benefits (i.e., when $w_A < .5$, and $w_B < .5$), conservative cutoff placement is predicted in the Figure 2b condition, but extreme cutoff placement is predicted in the Figure 2c condition. Figure 2d is similar to Figure 2b except that the numerical values in the first column (excluding the sign) were switched. Similarly, Figure 2e is like Figure 2b, except that the numerical values in the second column (excluding the sign) were switched. For Figure 2d, conservative cutoff placement results when the cost of an incorrect response to category A is weighted less heavily than the benefit of a correct category "A" response, and when the costs of an incorrect response to category B is weighted more heavily than the benefit of a correct category "B" response (i.e., when $w_A > .5$, and $w_B < .5$). The opposite pattern holds in Figure 2e (i.e., conservative cutoff placement results when $w_A < .5$, and $w_B > .5$). A comparison of performance across these four conditions will provide a test of the DWH, and will allow us to examine how the different cost-benefit structures influence performance. If the DWH is correct, and observers weight costs more heavily than benefits, then conservative cutoff placement should be observed in some conditions, and extreme cutoff placement should be observed in other conditions.

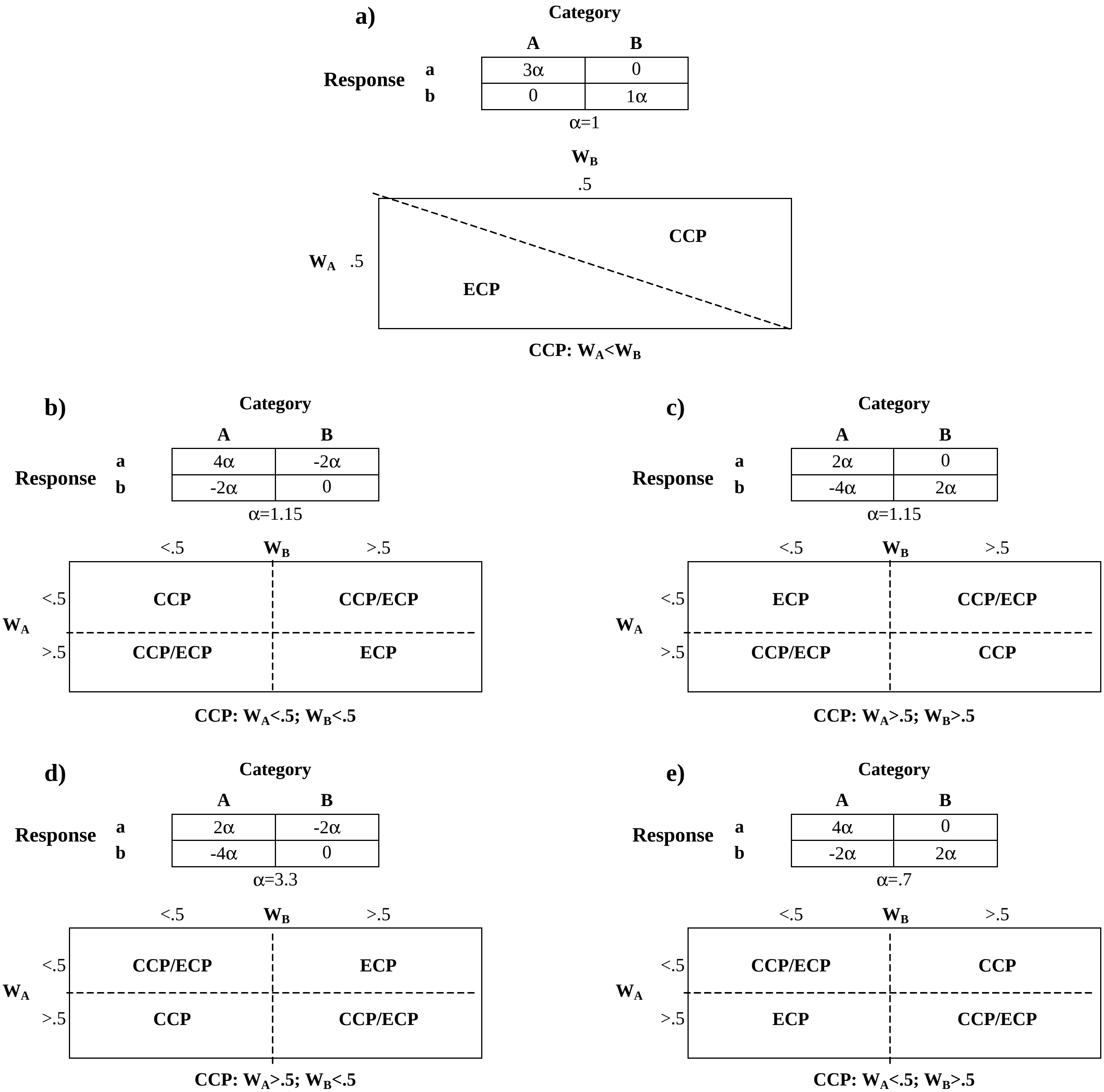

Figure 2. Cost-Benefit Structures and Differential Weighting Hypothesis predictions for various combinations of the weight values, w_A and w_B, for the zero cost condition (panel a), and for the four non-zero cost conditions (panels b – e).

A reasonable approach, and one we have taken in the past (Maddox, 1995; Maddox & Bohil, 1998a, 1998b, Experiment 1) is to utilize a between-observer design, apply the models to the individual observer data, and make comparisons across conditions. This approach would allow us to determine whether observers respond optimally, sub-optimally, or show payoff neglect. This approach also allows us to test the DWH by examining the β estimates from the S-O Payoff model. More recently, though, we have adopted a within-observer approach. This approach has several advantages. First, it minimizes betweengroup variability which increases statistical power. Second, it allows us to test all of the hypotheses described above, as well as to test a number of other hypotheses about possible relations among the observer's decision criteria across experimental conditions. This goal is achieved by developing nested decision bound models, each of which is applied to the data from all five experimental conditions *simultaneously* from a single observer. Each model instantiates a different set of hypotheses about the relations among the observer's decision criteria in each condition. For example, suppose we apply the optimal, sub-optimal and payoff neglect decision bound models separately to each experimental condition for a single observer, and find conservative cutoff placement in each condition. This is an interesting finding in its own right, but it would also be useful to determine whether the "conservative" decision criterion is identical in each condition (i.e., is unaffected by the presence of non-zero costs), or is different for zero and non-zero cost conditions. It is straightforward to develop models to test these and other hypotheses at the level of the individual observer. Finally, a within-observer design allows us to reduce the number of free parameters. For example, whereas separate fits of the five experimental conditions with the decision bound models would require five criterial noise parameters, simultaneous fits can be obtained in which the criterial noise is assumed to be constant across all five conditions (an assumption that is reasonable in the present case). In this article we used both approaches. First, we modeled the data from each condition separately (Separate Model Fits), then we applied a series of models simultaneously to the data from all five experimental conditions (Simultaneous Model Fits). The procedures for the separate fits have already been outlined. We detail the simultaneous model fit procedures in the section devoted to the Results and Theoretical Analyses.

All of the models tested in this article assume that the observer had knowledge of the category structures, and used the optimal decision function [i.e., $l(x_{pi}) = l_o(x_{pi})$]. This was an important assumption because we were interested in studying observers' sensitivity to cost and benefit manipulations, and not potential sub-optimalities in category distribution knowledge. To ensure that this was a reasonable assumption, the first session of the experiment was a baseline condition in which no cost-benefit manipulation was present. In addition, all observers completed a minimum of 60 baseline trials at the beginning of each of the five experimental conditions to ensure accurate knowledge of the category structures, and to minimize any within-observer carry-over effects from one experimental condition to the next. If the observer obtained a pre-determined performance criterion of 84% (only 2% below optimal) during the initial 60 baseline trials, they were allowed to begin the experimental trials. If the performance criterion was not met, they continued receiving baseline trials until the performance criterion was reached.

Method

Observers. Eight observers were solicited from the University of Texas community. All observers claimed to have 20/20 vision or vision corrected to 20/20. Observers were paid based on their day-to-day performance in the task.

Stimuli and Stimulus Generation. The stimulus was a horizontally varying dot that was presented within a 500 X 500 pixel square that was centered on a 1024 X 768 resolution computer screen. The vertical position of the dot remained unchanged, and the dot was centered vertically on the screen. There were two categories, A and B, each defined by a specific univariate normal distribution (Ashby & Gott, 1988). The Category A and B means were 45 pixels to the left and right of the center of the square, respectively, and the standard deviation for each category was 42 pixels. A representative stimulus from each category is depicted in Figure 1b. The category means and standard deviations resulted in a category level d' of 2.155. This values was chosen because previous research suggests that a category level d' in the range of 2 to 3 yields a fairly steep function relating the objective reward to the decision criterion (this is called the objective reward function; von Winterfeld & Edwards, 1982). Some have argued that shallow objective reward functions may adversely affect the optimality of observers' performance (Bohil & Maddox, 1999; Busemeyer & Myung, 1992; Maddox & Bohil, 1998b). The goal of the present research was to use a very steep objective reward function to improve observers' chances of responding optimally.

Prior to the experiment, a set of 60 stimuli was sampled randomly from each category distribution, yielding a total of 120 unique stimuli. In each block of 120 trials, every stimulus was presented once. Each session consisted of five blocks of trials in which the presentation order was randomized within each block. The first four blocks denoted the training phase. Trial-by-trial feedback was provided during the training phase (see details below). The final block of trials comprised the transfer phase. No feedback was provided during the transfer phase.

In the baseline condition the cost of an incorrect response was zero (i.e., $V_{aB} = V_{bA} = 0$), and the benefit of a correct response was 2 points (i.e., $V_{aA} = V_{bB} = 2$). The costs and benefits associated with A and B responses in each of the five experimental conditions are displayed in Figure 2. In all conditions, each point earned the observer $0.005. The point totals predicted by the optimal classifier (i.e., the classifier that maximizes long-run reward) for a block of 120 trials were 206, 213, 213, 213, 216, 216 in the baseline, and Figure 2a –2e experimental conditions, respectively. The optimal accuracy was 86% in the baseline condition, and was 83% in all five experimental conditions. In the baseline condition, $\beta_o = 1$, and in all experimental conditions, $\beta_o = 3$.

Procedure. Observers were told that perfect performance was impossible. However, an optimal level of performance was specified as a goal (in the form of desired point totals). Observers were informed that the dot would vary only in horizontal position, and that their goal was to maximize points in each session. They were informed that these point totals would be converted into monetary values that they would receive at the end of the experiment. On each day of the experiment, observers were given a 3-by-5 note card depicting the cost-benefit structure in effect for that day's condition. Observers were instructed to maximize points, and not worry about speed of responding. A typical trial proceeded as follows. A stimulus was presented on the screen, and remained until a response was made. The observer's task was to classify the presented stimulus as a member of Category A or Category B by pressing the appropriate button. During the training phase, the observer's response was followed by 750ms of feedback. Three lines of feedback were presented. The top line indicated the number of points the observer earned for the response. The next line indicated the potential point earnings for a correct response on each trial (i.e., if an observer responded incorrectly, this line indicated the amount that could have been earned had they chosen the correct response). The third line indicated the amount of money that the observer had accumulated up to that point in the session. The feedback was followed by a 125ms inter-trial interval in which the screen was blank. During the transfer phase, the observer's response was followed by an 875ms inter-trial interval in which the screen was blank. Observers were given periodic breaks during the training and transfer trials. At each break, the observer's accumulated point total was displayed.

During the first session, each observer completed five blocks in the baseline condition. The baseline condition was completed first to ensure that each observer had accurate knowledge of the category structures before being exposed to the cost-benefit manipulations. The five experimental conditions were completed, one session per day, on subsequent days starting with the Figure 2a condition and concluding with the Figure 2e condition. Before each experimental session, the observer completed a minimum of 60 baseline trials. If the observer reached a criterion of 84% correct (2% below optimal), then they were allowed to begin the experimental condition. If the observer did not reach criterion, they continued in the baseline condition until criterion was reached. Once they reached criterion, they were allowed to begin the experimental condition. Including these baseline trials prior to each experimental condition ensured that each observer had accurate knowledge of the category structures before exposure to the cost-benefit manipulation, and minimized the possibility of within-observer carry-over effects from one experimental condition to the next. The first experimental condition for all observers was the Zero cost condition (Figure 2a). For half of the observers, Category A was the high payoff category, and for half of the observers, Category B was the high payoff category. The high payoff category then alternated between Category A and B across the subsequent conditions. The "A" and "B" response buttons were counterbalanced across observers. To simplify the analyses, all data were reorganized so that Category A was the high payoff category.

Results and Theoretical Analysis

All model-based and other analyses were performed on the four blocks of training data, and on the transfer data. For completeness, and to facilitate future model testing, hit and false alarm rates (relative to the high payoff category) for each observer by block and condition are presented in Tables A-1 and A-2. We begin with a discussion of baseline condition performance and provide evidence that observers were able to learn the category distributions and use the optimal decision rule. We then turn to a discussion of the experimental conditions.

Baseline Condition

The baseline condition was included to ensure that observers had accurate knowledge of the category distributions prior to any cost-benefit manipulations. Two approaches were taken to determine whether we achieved this goal. First, we examined the accuracy rates, point totals, and the β estimates from the S-O Payoff model across the five blocks of baseline trials. Averaged across observers, the accuracy rates, point totals, and β estimates for each block are displayed in Figures 3a – 3c, respectively. Although not monotonic, performance approached that of the optimal classifier across blocks, and was nearly optimal during the final block of trials. Second, we fit the OPT, PON, and S-O Payoff models to the transfer block (block 5). For 6 of the 8 observers, the OPT model provided the most parsimonious account of the transfer data. For the remaining two observers the sub-optimal model provided the most parsimonious account of the transfer data, however the OPT model provided the most parsimonious account of the fourth block of training for these two observers. Restricting attention to the (averaged) S-O Payoff model β estimates, they steadily decreased from 1.15 during the first block to 1.028 by the final block of trials. In addition, the (averaged) OPT model fit accounted for 95.3% of the responses during the transfer block. Finally, averaged across observers, 85.2% accuracy (optimal = 86%), and 204.5 points (optimal = 206 points) were obtained during the transfer phase. Taken together, the model-based analyses, accuracy rates, and point totals suggest that observers learned the category distributions, and used the optimal decision rule by the end of the baseline condition.

Experimental Conditions

Summary Statistics and S-O Payoff Model Fits By Block

Tables A-1 and A-2 display the hit rate and false alarm rates (where a hit was defined as a correct category “A” response, and a false alarm was defined as an incorrect category “B” response to a Category A exemplar) for each of the five Figure 2 conditions by observer and block. The (averaged) accuracy rates (derived from the hit and false alarm data), point totals, and S-O Payoff model β estimates are displayed in Figures 4a – 4c, respectively. [The β value for Observer 5, block 3, condition 2b was removed from Figure 4c because it was several standard deviations removed from the rest of the data.] The thick solid line denotes the Figure 2a, zero cost condition data. The data averaged across the four non-zero cost conditions are also included in each figure and are denoted by the thick broken line. This facilitates a comparison of the zero and non-zero cost conditions. The data are variable across blocks, but a focus on the first and last training blocks (Blocks 1 and 4) and the transfer block suggests several interesting findings. First, there was a small increase in the point totals across blocks with a slight advantage for the zero cost over the (averaged) non-zero cost condition. The true zero vs. non-zero cost difference is potentially larger as the optimal point total was slightly higher in the Figure 2d and 2e conditions--both non-zero cost conditions. Second, accuracy showed a general decline from the first to the last training block, but then increased from the last training block to the transfer block. [Recall that the optimal classifier maximizes long-run reward, and thus, when costs and benefits are manipulated, must sacrifice accuracy to attain this goal.] In addition, during the final training block and during the transfer block, accuracy in the zero cost condition was lower than in any of the non-zero cost conditions, suggesting that observers were more willing to sacrifice accuracy when there were no costs associated with an incorrect response. Third, the S-O Payoff model β estimates showed a general increase from the first to the last training block, and were closer to the optimal value ($\beta_o = 3$) in the zero cost condition. Interestingly, though, the β value continued approaching the optimal value during the transfer block for the zero cost condition, but became less optimal (i.e., decreased toward 1) during the transfer block in the non-zero cost conditions. Finally, there was little support for the DWH. The strongest test of the DWH requires a comparison of the β estimates from the Figure 2b and 2c conditions. Specifically, if the DWH is supported, then conservative cutoff placement should result in the Figure 2b condition, and extreme cutoff placement should result in the Figure 2c condition. This was not the case; conservative cutoff placement was prevalent in both conditions. Although these data do not support a strong version of the DWH, a weaker version of the DWH can be formulated. Whereas the strong version

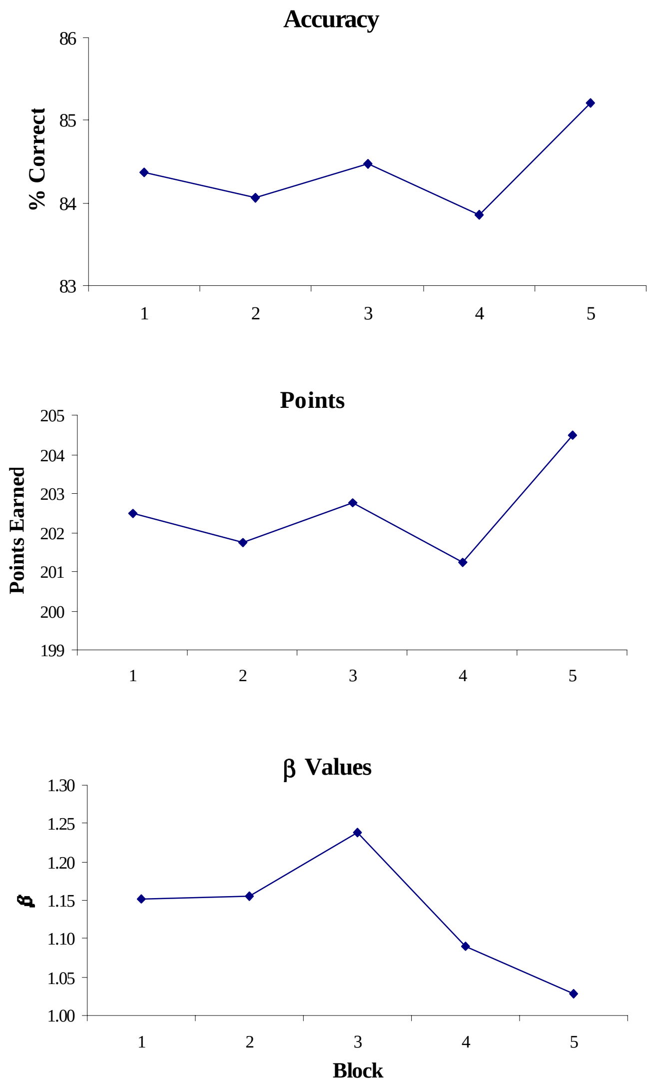

Figure 3. Accuracy rates, point totals and S-O Payoff model decision criterion from the baseline condition by block averaged across observers.

requires extreme cutoff placement in Figure 2c, and conservative cutoff placement in Figure 2b, a weaker version would require only that the decision criterion be larger in Figure 2c than in Figure 2b. This version received some support. During the transfer block, the weak version of the DWH was supported for 5 of the 8 observers, but was supported for only 2 of the 8 observers during the final training block.

At this stage, a few comments are in order regarding the qualitative shift in performance across the training and transfer blocks. Within the framework of the S-O Payoff model there are two factors that affect accuracy. One is the placement of the decision criterion β. In the current study, as β approaches the optimal value, accuracy should decrease. Thus, it is not surprising that accuracy decreased during training while the β estimate increased. The second is the trial-by-trial variability in the placement of the decision criterion (termed criterial noise). In general, the larger the criterial noise (for a fixed decision criterion), the lower the accuracy. The criterial noise estimates from the S-O Payoff model showed a moderate (although non-monotonic) decline across the training blocks

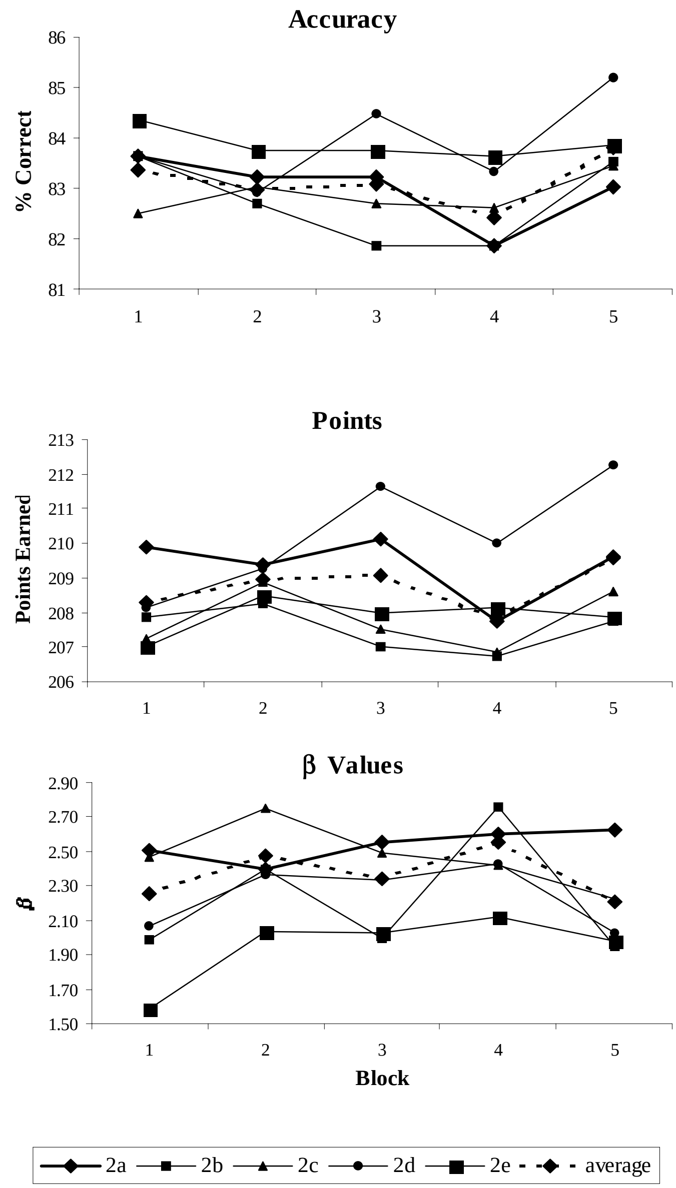

Figure 4. Accuracy rates, point totals and S-O Payoff model decision criterion from the five experimental conditions (as well as the average of the four non-zero cost conditions) by block averaged across observers.

followed by a sharp decline during the transfer block[3]. Observers were instructed to use a fixed decision criterion during the transfer block, and the noise estimates suggest that observers followed this instruction. One possibility is that the increase in accuracy during the transfer block for the zero cost condition was due to the reduction in criterial noise, and not to a decrease in the β value since the β estimates increased. In the non-zero cost condition on the other hand, the increase in accuracy was likely due to both a decrease in criterial noise, and a decrease in the β value associated with a greater emphasis on accuracy. Although speculative, it appears that the withdrawal of feedback (during the transfer trials) led to a greater emphasis on accuracy when non-zero costs were incurred, but to a continued emphasis on reward when the cost of an incorrect response was zero. Clearly further research is needed to test this speculative hypothesis. Even so, the remaining model-based analyses will be performed on the final training block, and on the transfer block.

Separate Model Fits of the Final Training Block, and the Transfer Block

Tables 1a and 1b present the most parsimonious model, percent of responses accounted for by the most parsimonious model and the β estimates from the most parsimonious model by observer and condition for the separate model fits. Table 1a presents the data for the final training block, and Table 1b presents the data for the transfer block. Several interesting results emerge. First, there was little evidence of extreme cutoff placement. In fact, in only 6 of the 40 cases (8 observers X 5 cost-benefit structures) during the final block of training, and only 2 of 40 cases during transfer was extreme cutoff placement observed, and 6 of these were shown for the same Observer (Observer 6). Second, the remaining cases were fairly evenly distributed between optimal cutoff placement (final block of training: 14 of 40; transfer: 11 of 40), conservative cutoff placement (final block of training: 10 of 40; transfer: 14 of 40), and payoff neglect (final block of training: 10 of 40; transfer: 13 of 40). Third, the average β value was closer to the optimal value in the zero cost condition than in the non-zero cost conditions during the final block of training and during transfer. Finally, based on the β values from the most parsimonious model we again find some support for the weaker version of the DWH, but not for the stronger version. Specifically, the averaged β values are larger in the Figure 2c condition than in the Figure 2b condition, but only during the transfer block. During the final block of training the opposite holds. We turn now to the simultaneous model fits.

Simultaneous Model Fits of the Tranfer Data for All Experimental Conditions

The nested structure of the simultaneous fit models is outlined in Figure 5. The arrows point to a more general model. Two 1-parameter models were tested. The *optimal decision bound model* (OPT) assumes that the observer's decision criterion was optimal in each condition, and the observer attempted to maximize long-run reward. The only free parameter was the noise parameter. The *payoff neglect model* (PON) assumes that the observer was completely insensitive to, and neglected, the payoff differences, which is equivalent to maximizing categorization accuracy. Again the only free parameter was the noise parameter. Two 2-parameter models were tested. The *sub-optimal payoff model* (S-O Payoff) assumes that the observer was sensitive to the cost-benefit manipulation, but used a sub-optimal decision criterion that was identical in all experimental conditions. This model instantiates the hypothesis that performance is unaffected by the presence (or absence) of non-zero costs. The *sub-optimal zero cost; payoff neglect non-zero cost model* (S-O, Zero; PON, Non-zero) assumes that the observer was sensitive to the cost-benefit manipulation when the costs were zero, but showed payoff neglect when the costs were non-zero. The idea here is that the presence of non-zero costs leads the observer to maximize accuracy and sacrifice reward. One 3-parameter model was tested. The *sub-optimal zero cost; sub-optimal non-zero cost model* (S-O Zero; S-O Non-zero) assumes that the observer was sensitive to the cost-benefit manipulation, but allows the zero cost decision criterion to be different from the non-zero cost decision criterion, all of which are assumed to be equal. This model instantiates the hypothesis that non-zero costs have a different effect on decision criterion placement than zero costs. However, this model makes the simple assumption that the presence

[3] As stated earlier, the noise parameter represents the sum of perceptual and criterial noise, and thus does not provide an independent estimate of criterial noise. However, since the stimuli were of high contrast and the display was response terminated, it is likely that the majority of this estimate is due to the effects of criterial noise.

Table 1a.

Most Parsimonious Model, Percent of Responses Accounted for by Most Parsimonious Model, and β Estimate From the Most Parsimonious Model by Condition and Observer for the Final Training Block

Observer	Condition	Most Parsimonious Model	% of Responses Accounted For	β Values
	2a	**S-O PAY**	97.50	2.00
	2b	**S-O PAY**	98.33	1.64
1	2c	**PON**	93.33	1.00
	2d	**S-O PAY**	98.33	1.81
	2e	**S-O PAY**	97.50	0.68
	2a	**OPT**	92.50	3.00
	2b	**PON**	96.67	1.00
2	2c	**S-O PAY**	95.00	1.97
	2d	**OPT**	96.67	3.00
	2e	**S-O PAY**	95.83	1.53
	2a	**PON**	90.00	1.00
	2b	**OPT**	91.67	3.00
3	2c	**PON**	94.17	1.00
	2d	**S-O PAY**	92.50	1.72
	2e	**PON**	87.50	1.00
	2a	**OPT**	92.50	3.00
	2b	**PON**	87.50	1.00
4	2c	**PON**	95.00	1.00
	2d	**OPT**	91.67	3.00
	2e	**PON**	91.67	1.00
	2a	**PON**	85.00	1.00
	2b	**S-O PAY**	95.83	1.66
5	2c	**S-O PAY**	97.50	1.51
	2d	**PON**	95.83	1.00
	2e	**S-O PAY**	99.17	1.46
	2a	**S-O PAY**	94.17	5.60
	2b	**OPT**	94.17	3.00
6	2c	**S-O PAY**	94.17	4.35
	2d	**S-O PAY**	99.17	4.66
	2e	**S-O PAY**	94.17	4.66
	2a	**OPT**	95.00	3.00
	2b	**S-O PAY**	88.33	7.06
7	2c	**S-O PAY**	87.50	5.88
	2d	**OPT**	90.83	3.00
	2e	**OPT**	96.67	3.00

(table continues)

Table 1a (continued)

8	2a	**OPT**	91.67	3.00
	2b	**OPT**	94.17	3.00
	2c	**OPT**	96.67	3.00
	2d	**OPT**	95.83	3.00
	2e	**OPT**	98.33	3.00
average	2a		92.29	2.70
	2b		93.33	2.67
	2c		94.17	2.46
	2d		95.10	2.65
	2e		95.10	2.04

Table 1b.

Most Parsimonious Model, Percent of Responses Accounted for by Most Parsimonious Model, and β Estimate From the Most Parsimonious Model by Condition and Observer for the Transfer Block

Observer	Condition	Most Parsimonious Model	% of Responses Accounted For	β Values
1	2a	**S-O PAY**	100.00	1.54
	2b	**PON**	99.17	1.00
	2c	**S-O PAY**	95.00	2.19
	2d	**S-O PAY**	95.83	1.82
	2e	**PON**	97.50	1.00
2	2a	**PON**	95.00	1.00
	2b	**S-O PAY**	98.33	1.58
	2c	**S-O PAY**	98.33	1.97
	2d	**S-O PAY**	95.00	1.99
	2e	**S-O PAY**	96.67	1.83
3	2a	**S-O PAY**	95.83	2.09
	2b	**PON**	90.83	1.00
	2c	**PON**	95.00	1.00
	2d	**PON**	95.00	1.00
	2e	**PON**	95.00	1.00
4	2a	**S-O PAY**	92.50	1.75
	2b	**PON**	82.50	1.00
	2c	**PON**	89.17	1.00
	2d	**S-O PAY**	88.33	1.80
	2e	**PON**	92.50	1.00
5	2a	**PON**	95.00	1.00
	2b	**S-O PAY**	97.50	1.27
	2c	**S-O PAY**	96.67	1.63
	2d	**PON**	94.17	1.00
	2e	**PON**	97.50	1.00

(table continues)

Table 1b (continued)

	2a	**S-O PAY**	98.33	6.65
	2b	**OPT**	95.83	3.00
6	2c	**OPT**	98.33	3.00
	2d	**OPT**	100.00	3.00
	2e	**S-O PAY**	97.50	4.16
	2a	**OPT**	94.17	3.00
	2b	**OPT**	97.50	3.00
7	2c	**OPT**	90.00	3.00
	2d	**S-O PAY**	95.83	2.09
	2e	**OPT**	95.00	3.00
	2a	**OPT**	96.67	3.00
	2b	**OPT**	95.00	3.00
8	2c	**S-O PAY**	98.33	2.53
	2d	**OPT**	98.33	3.00
	2e	**OPT**	96.67	3.00
	2a		95.94	2.50
	2b		94.58	1.86
average	2c		95.10	2.04
	2d		95.31	1.96
	2e		96.04	2.00

of any non-zero cost, regardless of its magnitude, leads to the same decision criterion. Three 4-parameter models were tested. Each of these models generalizes the sub-optimal zero cost; sub-optimal non-zero cost model by assuming that one pair of non-zero cost conditions yield one decision criterion, and that the other pair of non-zero cost conditions yield a different decision criterion. We will refer to these as the *sub-optimal zero cost; sub-optimal non-zero cost, differential pair models*. One version of this model assumes a single decision criterion in the Figure 2b and 2d conditions, and a different decision criterion in the Figure 2c and 2e conditions. Because this model pairs the non-zero cost conditions that have the same cost-benefit structure for the low payoff category (i.e., the "B" column of each cost-benefit structure), we refer to this as the *sub-optimal zero cost; sub-optimal non-zero cost, low payoff pairing model*. A second version of this model assumes a single decision criterion in the Figure 2b and 2e conditions, and a different decision criterion in the Figure 2c and 2d conditions. Because this model pairs the non-zero cost conditions that have the same cost-benefit structure for high payoff cost-benefit structures are switched, we refer to this as the *sub-optimal zero cost; sub-optimal non-zero cost, switched payoff pairing model.* Each of these 4-parameter models contradicts the strong the high payoff category (i.e., the "A" column of each cost-benefit structure), we refer to this as the *sub-optimal zero cost; sub-optimal non-zero cost, high payoff pairing model*. The third version of this model version of the DWH because each assumes the same assumes a single decision criterion in the Figure 2b and 2c conditions, and a different decision criterion in the Figure 2d and 2e conditions. Because this model pairs the non-zero conditions for which the low and decision criterion in conditions for which the strong version of the DWH would predict qualitatively different decision criteria (i.e., extreme cutoff placement in one case, and conservative cutoff placement in another case). One 6-parameter model was tested. The *general sub-optimal payoff model* assumes that the observer was sensitive to the cost-benefit manipulation, but this model allowed the decision criterion in each condition to differ from that in every other condition. If the strong version of the DWH is supported, then this model should provide the most parsimonious account of the data, and the

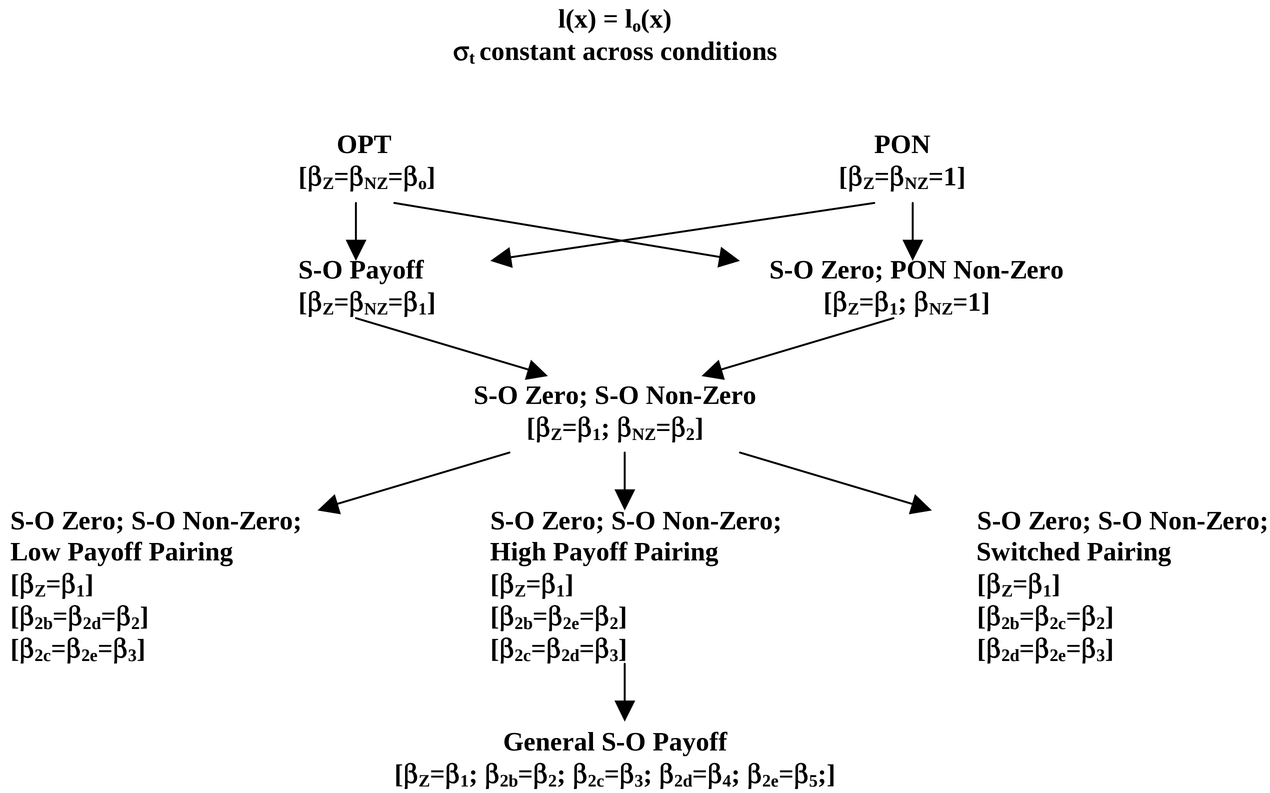

Figure 5. Nested relationship among the decision bound models applied simultaneously to the data from all five experimental conditions. The arrow points to a more general model.

resulting decision criterion estimates from each condition should be in line with those outlined in Figure 2.

Each of the Figure 5 models was applied to the data from each individual observer. The most parsimonious model, percent of responses accounted for by the most parsimonious model, and β estimates from the most parsimonious model by condition and observer are summarized in Tables 2a and 2b. Table 2a presents the data for the final training block, and Table 2b presents the data for the transfer block. Recall that the simultaneous modeling procedure allowed us to test several hypotheses about cost-benefit knowledge that we could not test with the separate model fits, and allowed us to accomplish this with fewer free parameters. However, this is an advantage only if the simultaneous modeling approach provides a nearly equivalent account of the data. We compared the overall performance of the simultaneous and separate modeling procedures by computing the percent of responses accounted for by the most parsimonious simultaneous and separate fit models. For both the final training and the transfer block, the percent of responses accounted for were 94% or higher, and never differed by more than 1% when comparing separate with simultaneous fits. Clearly the simultaneous models provided a good, and nearly equivalent, description of the data.

Several interesting results emerge from an examination of Tables 2a and 2b. First, 5 of the 8 observers (Observers 1–5) showed conservative cutoff placement in all five experimental conditions for both the final training (Table 2a) and transfer block (Table 2b). Two of these observers (Observers 3 and 5 during the final training block, and Observers 4 and 5 during the transfer block) treated the zero and non-zero cost conditions equivalently, and used the same sub-optimal decision criterion in all conditions (i.e., the most parsimonious model was the S-O Payoff model). The remaining three observers used different decision criterion in the zero and non-zero cost conditions. Second, Observer 8 used the same decision criterion in all conditions, and used a decision criterion that was close to optimal. In fact, during the final training block, the optimal decision bound model provided the most parsimonious account of this observer’s data. Third, Observers 6 and 7 showed a very different pattern of responding from the other 6 observers.

Table 2a.

Most Parsimonious Figure 5 Model, β estimates, and Percent of Responses Accounted for by Most Parsimonious Figure 5 Model for the Experimental Conditions from the Final Training Block.

		Experimental Condition					
		2a	2b	2c	2d	2e	
Observer	Most Parsimonious Model	β values					% accounted for
1	**S-O Zero; S-O Non-Zero Low Payoff Pairing**	1.94	1.60	0.90	1.60	0.90	96.3
2	**S-O Zero; S-O Non-Zero High Payoff Pairing**	2.51	1.36	2.15	2.15	1.36	95.3
3	**S-O Payoff**	1.51	1.51	1.51	1.51	1.51	91.3
4	**S-O Zero; S-O Non-Zero Low Payoff Pairing**	2.26	1.84	1.00	1.84	1.00	91.8
5	**S-O Payoff**	1.43	1.43	1.43	1.43	1.43	95.3
6	**S-O Zero; S-O Non-Zero**	5.63	4.15	4.15	4.15	4.15	94.5
7	**S-O Zero; S-O Non-Zero Switched Payoff Pairing**	2.84	5.91	5.91	2.67	2.67	92.2
8	**OPT**	3.00	3.00	3.00	3.00	3.00	95.3
Average		2.64	2.60	2.51	2.29	2.00	94.0
Optimal		3	3	3	3	3	

Note. % Accounted for = Percent of Responses Accounted for by Most Parsimonious Figure 5 Model.

Table 2b.

Most Parsimonious Figure 5 Model, β estimates, and Percent of Responses Accounted for by Most Parsimonious Figure 5 Model for the Experimental Conditions from the Transfer Block.

		Experimental Condition					
		2a	2b	2c	2d	2e	
Observer	Most Parsimonious Model	β values					% Accounted for
1	**S-O Zero; S-O Non-Zero High Payoff Pairing**	1.64	1.01	2.04	2.04	1.01	97.5
2	**S-O Zero; S-O Non-Zero**	1.31	1.87	1.87	1.87	1.87	96.7
3	**S-O Zero; PON Non-Zero**	1.97	1.00	1.00	1.00	1.00	94.5

(table continues)

Table 2b (continued)

4	**S-O Payoff**	1.30	1.30	1.30	1.30	1.30	89.3
5	**S-O Payoff**	1.31	1.31	1.31	1.31	1.31	96.0
6	**S-O Zero; S-O Non-Zero**	6.62	3.45	3.45	3.45	3.45	97.8
7	**S-O Zero; S-O Non-Zero Low Payoff Pairing**	4.01	2.46	3.71	2.46	3.71	94.8
8	**S-O Payoff**	2.54	2.54	2.54	2.54	2.54	97.0
Average		2.59	1.87	2.15	2.00	2.02	95.5
Optimal		3	3	3	3	3	

Note. % Accounted for = Percent of Responses Accounted for by Most Parsimonious Figure 5 Model.

Observer 6 showed extreme cutoff placement in all conditions, but a more extreme cutoff in the zero cost condition (Figure 2a). Perhaps this observer invoked a response bias toward the "high payoff" category (see Maddox & Bohil, 1998b, for a discussion of response biases in categorization). Observer 7's pattern of responding is unique in many ways, and is quite different during the final training block as compared with the transfer block. Most importantly, Observer 7 is the only observer to show support for the strong version of the DWH. Specifically, during the transfer block this observer showed extreme cutoff placement in the Figure 2c condition, and conservative cutoff placement in the Figure 2b condition. Unfortunately, the same pattern was not observed during the final training block. Finally, some support for the weak version of the DWH was again obtained. Specifically, the average β value for the Figure 2c condition was larger than for the Figure 2b condition, but only during the transfer block. Even so, this was due mainly to the results from Observer's 1 and 7. For the remaining observers, the β estimates from the most parsimonious model were identical in these two conditions.

General Discussion

This article reports the results of an experiment that attempts to bridge the gap between categorization and decision-making research by examining human categorization performance when the costs and benefits associated with each categorization response are manipulated. The cost-benefit structures differed across conditions, but were selected so that the performance of the optimal classifier (derived from the expected value rule) was identical across conditions. Each observer completed several blocks of trials with each of the cost-benefit structures. This allowed us to capture the time-course of cost-benefit learning, and provides a rich database for current and future model testing. One focus was to compare performance when the cost of an incorrect response was zero versus non-zero. A second focus was to provide a test of the hypothesis that observers weight costs more heavily than benefits during categorization. This hypothesis has received some support in the decision-making literature, but to date has not been tested in categorization.

In general, observers showed sensitivity to the cost-benefit structures in both the zero and non-zero cost conditions. As the observers gained experience with the cost-benefit structures, model-based estimates of the decision criterion and point totals increased toward the optimal values, and accuracy rates declined. Observers' performance was consistently better in the zero cost condition than in the non-zero cost conditions. In addition, during the transfer block, in which trial-by-trial feedback was removed, observers' continued to show improved performance in the zero cost condition, but show decrements in performance in the non-zero cost conditions. Across all conditions, the most common finding was conservative cutoff placement (the use of a decision criterion more conservative than that used by the optimal classifier). Finally, only modest support was found for the differential weighting hypothesis. We now turn to a more detailed discussion of the more salient findings.

Prevalence of Conservative Cutoff Placement

Across a number of studies that focused on base-rate and/or category benefit learning, the most common finding was that observers were sensitive to the manipulation, but that they tended to use a decision criterion that was more conservative than that predicted by the optimal classifier (Busemeyer & Myung, 1992; Bohil & Maddox, 1999; Maddox & Bohil, 1998a, 1998b). This is commonly referred to as conservative cutoff placement. The current study included several conditions in which both the costs and benefits were manipulated. Again the most prevalent finding was conservative cutoff placement. Several explanations for conservative cutoff placement have been offered in the literature. We briefly review these and offer another explanation.

Distribution Misconception Hypothesis

One hypothesis is that conservative cutoff placement is caused by a systematic misconception of the shape of the underlying distributions (e.g., Kubovy, 1977; Kubovy & Healy, 1980; Maloney & Thomas, 1991). This hypothesis predicts correctly the use of a conservative decision criterion, but fails on two accounts. First, Healy and Kubovy (1981) and Maddox and Bohil (1998b) provide evidence that responding is more nearly optimal when base-rates, as opposed to costs and benefits, are manipulated. To account for this finding, the systematic misconception hypothesis would have to predict different misconceptions when base-rates versus costs and benefits are manipulated. It is unclear why decisional variables, such as base-rates and costs and benefits, would differentially affect knowledge of the distributions (however, see Kruschke, 1996). Second, Maddox and Bohil (1998b; see also Busemeyer & Myung, 1992) provided preliminary evidence that the category-level d’ also affects the magnitude of conservative cutoff placement. Specifically, d’ values near 1 yield conservative cutoff placement of a large magnitude, whereas d’ values near 2 yield conservative cutoff placement of a smaller magnitude. Again, it is unclear how the category-level d’ would affect the perceived shape of the category distributions, when d’ can be manipulated without affecting the shape of the distributions.

Generalized Probability-Matching Hypothesis

Another hypothesis that has received some discussion is the generalized probability-matching hypothesis (Healy & Kubovy, 1981; Thomas & Legge, 1970). In short, it assumes that the observer selects a decision criterion that matches their response probabilities to the base-rates plus a constant where the constant is determined from the cost-benefit structure. T-tests were performed to determine whether the observed P(“A”) in each of the five experimental conditions were significantly larger than .50. For the final training block these values, averaged across observers, were .594, .585, .572, .561, and .583 in the Figure 2a – 2e conditions, respectively. For the transfer block these values, in the same order and again averaged across observers, were .586, .560, .570, .555, and .565. All 10 t-tests were significant at the .05 level providing evidence for the generalized probability-matching hypothesis. To determine whether the constant from the generalized probability-matching hypothesis was affected by the presence of non-zero costs, we performed one-way ANOVAs on the observed P(“A”) for the five experimental conditions separately for the final training and transfer blocks. The ANOVAs were non-significant [final training block: $F(4, 28) = .966$, $p > .40$; transfer block: $F(4, 28) = 1.427$, $p > .20$] suggesting that the constant predicted by the generalized probability-matching hypothesis is unaffected by the presence of non-zero costs.

It is worth mentioning that the generalized probability-matching hypothesis is non-committal about the mechanism used to construct the decision criterion, it assumes only that the decision criterion is selected so that the P(“A”) is equal to the base-rate plus a constant. Thus, the hypothesis is not constrained to predict that the observer uses a strategy similar in spirit to that of the optimal classifier (i.e., compute the ratio of the difference between the costs and benefits). In fact, because probability matching is often non-optimal, it is likely that the generalized probability-matching hypothesis would involve a very different set of processes underlying construction of the decision criterion. Decision bound theory assumes that observers use the same strategy as the optimal classifier (albeit in the presence of various potential sub-optimalities, such as perceptual and criterial noise), and so is fundamentally at odds with the generalized probability-matching hypothesis. Because the generalized probability-matching hypothesis assumes that the observed P(“A”) is affected only by the base-rates and the cost-benefit structure, other factors such as category-level d’ should have no effect. The optimal classifier, on the other hand, is strongly affected by the category-level d’. For example, when d’ = 2.155 (as in the current study), the optimal classifier will respond “A” with probability .62. However, when d’ = 1, the optimal classifier will respond “A” with a larger probability equal to .83. Healy and Kubovy (1981) conducted a study that was similar in spirit to ours, and included a condition like our zero cost condition except that the category-level d’ = 1. In their zero cost condition, they observed a P(“A”) = .665. Our study used a d’ = 2.155, and in the zero cost condition we observed P(“A”) = .586. Unfortunately, Healy and Kubovy (1981) did not report the probabilities for individual observers, so we are unable to determine whether this difference is statistically significant[4]. Although tentative at this point, we feel that the generalized probability-matching hypothesis is insufficient because it predicts no effect of category d’. Kubovy and Healy (1980) briefly reported the results of a pilot study in which d’ was manipulated, and found that d’ affected the β estimates [and thus the P(“A”)] in a way predicted by the optimal classifier. Future research should address the relationship between category-level d’ and the optimality of performance.

Competition Between Reward and Accuracy (COBRA)

Maddox and Bohil (1998b) recently proposed an alternative to the generalized probability-matching hypothesis that accounts for the prevalence of conservative cutoff placement, and the effects of d’ on P(“A”) when costs and benefits (and/or base-rates) are manipulated. Unlike the generalized probability-matching hypothesis, Maddox and Bohil’s proposal assumes that the observer uses the same strategy as the optimal classifier, but that there is a COmpetition Between Reward and Accuracy (COBRA) maximization. The idea is that observers attempt to maximize reward, as instructed, but also place some importance on the accuracy of their responding. When costs and benefits are manipulated, as in the present study, both goals cannot be achieved simultaneously because the decision rule that maximizes accuracy is different from the decision rule that maximizes reward (see Figure 1a). An observer who places importance on both goals will use an intermediate decision criterion, and will show conservative cutoff placement. With experience, the observer might learn to adjust the weighting function in favor of reward maximization which would lead to a shift in the observed decision criterion toward that of the optimal value. Even so, as long as some importance is assigned to accuracy maximization, conservative cutoff placement will be observed. In addition, because COBRA is driven by the goals of reward and accuracy maximization, the observed probability of responding “A” will generally be affected by the probability of responding “A” predicted by each goal, and thus should be affected by such factors as d’. The notion that there is a competition, or trade-off, between the accuracy of categorization responses and long-run reward likely affects nearly all categorization problems. For example, a medical doctor faces two (potentially competing) goals. One is to accurately diagnose each patient. The second is to achieve the first without adversely affecting the profit margin. The profit margin is driven by the goal to maximize long-run reward, whereas accurate diagnosis is driven by the goal to maximize long-run accuracy. When the pattern of symptoms is such that the correct diagnosis is obvious, few costly tests are needed and high levels of accuracy and reward can be achieved simultaneously. On the other hand, when costly tests are necessary to secure an accurate diagnosis, it is possible that long-run reward may need to be sacrificed. The rise in health care costs and the proliferation of Health Maintenance Organizations (HMO’s) have increased the likelihood of this sort of accuracy-reward competition in medical decision making.

Effects of Non-Zero Costs

A major finding of the current study was that observers were closer to optimal in the zero cost than in the non-zero cost conditions, even though the performance predicted by the optimal classifier was identical across all conditions. One possibility is that the objective reward function was steeper in the zero

[4] The reported value of .665 is only an estimate. Healy and Kubovy (1981) presented these data in the form of a figure, and so exact values were unavailable.

cost, than in the non-zero cost conditions. Although the category-level d' was chosen to maximize the steepness of the objective reward function, this holds only relative to a fixed cost-benefit structure. Because the cost-benefit structures differ across conditions, the objective reward functions differ, although they all peak at the same β value with the same maximum objective reward. Interestingly, the objective reward function for the zero-cost condition is the shallowest, not the steepest. Thus, if performance is driven solely by the steepness of the objective reward function, then performance should be worst, not best, in the zero-cost condition. Bohil and Maddox (1999; see also Maddox & Bohil, 1998b) provide evidence in support of the claim that steeper objective reward functions lead to superior performance, but the present study suggests the nature of the costs has a separate, perhaps independent, effect.

An explanation that we feel has merit is that the presence of non-zero costs leads the observer toward a greater emphasis on accuracy. Within the framework of COBRA, the weight assigned to accuracy maximization would be larger when non-zero costs are present. This is a sort of differential weighting hypothesis, but instead of postulating differential weight to the costs over the benefits, we postulate differential weight to accuracy maximization when costs are non-zero. There are at least two aspects of our data that support this claim. First, performance was worse (in terms of reward maximization) in the non-zero cost conditions. Specifically, β values and point totals were lower, and accuracy was higher. Second, performance declined in the non-zero cost conditions when feedback was removed, but continued to improve in the zero cost condition. Clearly this hypothesis is tentative and awaits more rigorous testing. Even so, it does provide a reasonable account of the qualitative aspects of the data.

Implications for Decision Theory

A major goal of this research was to bridge the gap between decision-making and categorization. We have made some inroads toward this goal, but primarily focused our discussion on categorization. Even so, our hope is that decision theorists will take advantage of the database offered by this study, and detailed in the Appendix, and will begin to apply their models to categorization data. Promising approaches are offered by Busemeyer and Myung's (1992) Hill-Climbing model, Erev's (1998; see also Erev, Wallsten, & Budescu, 1994) Cutoff Reinforcement Learning Model, and Wallsten and Gonzalez-Vallejo's (1994) Stochastic Judgment Model. One possibility would be to use the Hill-Climbing or Cutoff Reinforcement Learning models as a learning mechanism for COBRA.

In conclusion, the present study extends our understanding of the use of cost-benefit knowledge in perceptual categorization to several situations in which the costs of incorrect responses were zero or non-zero. The cost-benefit structures were constrained in such a way that the performance of the optimal classifier was identical in each condition. In general, performance became more nearly optimal as the observers gained experience with the cost-benefit structures, but performance was consistently worse when non-zero costs were present. One possibility is that observers placed some importance on accuracy maximization as well as reward maximization, and that the presence of non-zero costs increased the emphasis on accuracy maximization.

References

Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449-484). Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 33-53.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General, 120*, 150-172.

Ashby, F. G., & Lee, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In S. C. Masin (Ed.), *Foundations of perceptual theory* (pp. 369-397). New York: North Holland.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology, 37,* 372-400.

Ashby, F. G., & Maddox, W. T. (1994). A response time theory of perceptual separability and perceptual integrality in speeded classification. *Journal of Mathematical Psychology, 33*, 423-466.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154-179.

Bohil, C. J., & Maddox, W. T. (1999). Base-rates, payoffs, and category-level discriminability. *Manuscript in preparation*.

Busemeyer, J.P. & Myung, I.J. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General*, *121*, 177-194.

Erev, I. (1998). Signal detection by human observers: A cutoff reinforcement learning model of categorization decisions under uncertainty. *Psychological Review*, *105*, 280-298.

Erev, I., Wallsten, T. S., & Budescu, D. V., (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101,* 519-527.

Estes, W. K., Campbell, J. A., Hatsopoulos, N., & Hurwitz, J. B. (1989). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 556-571.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*, NewYork: Wiley.

Healy, A. F., & Kubovy, M. (1981). Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 344-354.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*, 319-340.

Holyoak, K. J., & Spellman, B. A. (1993). Thinking. *Annual Review of Psychology*, *44*, 265-315.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237-251.

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

Koehler, J. J. (1996). The base-rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19,* 1-53.

Kruschke, J. K. (1996). Base-rates in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 3-26.

Kubovy, M. (1977). A possible basis for conservatism in signal detection and probabilistic categorization tasks. *Perception & Psychophysics, 22,* 277-281.

Kubovy, M., & Healy, A. F. (1980). Process models of probabilistic categorization. In T. S. Wallsten (Ed.), *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, N. J.: Erlbaum.

Macmillan, N.A., & Creelman, C.D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Maddox, W. T. (1995). Base-rate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 288-301.

Maddox, W.T. (1999). On the danger of averaging across observers when comparing decision bound and generalized context models of categorization. *Perception & Psychophysics, 61,* 354-374.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics, 53,* 49-70.

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification - speeded classification relationship. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 795-817.

Maddox, W. T., & Ashby, F. G. (1998). Selective attention and the formation of linear decision boundaries: Comment on McKinley and Nosofsky (1996). *Journal of Experimental Psychology: Human Perception and Performance, 24*, 301-321.

Maddox, W. T., & Bogdanov, S.V. (1999). On the relation between decision rules and perceptual representation in multidimensional perceptual categorization. *Manuscript submitted for publication*.

Maddox, W. T., & Bohil, C. J. (1998a). Overestimation of base-rate differences in complex perceptual categories. *Perception and Psychophysics, 60,* 575-592.

Maddox, W. T., & Bohil, C. J. (1998b). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 1459-1482.

Maloney, L. T., & Thomas, E. A. C. (1991). Distributional assumptions and observed conservatism in the theory of signal detectability. *Journal of Mathematical Psychology*, *35*, 443-470.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.

Stevenson, M. K., Busemeyer, J. R., & Naylor, J.C. (1991). Judgment and decision-making theory. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology, Vol. 1 (2nd ed.)*, pp. 283-374.

Thomas, E. A. C. & Legge, D. (1970). Probability matching as a basis for detection and recognition decisions. *Psychological Review*, *77*, 65-72.

Thomas, R. D. (1995). Gaussian general recognition theory and perceptual independence. *Psychological Review*, *102*, 192-200.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124-1131.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in Social Psychology*. Erlbaum.

von Winterfeldt, D., & Edwards, W. (1982). Costs and payoffs in perceptual research. *Psychological Bulletin, 91*, 609-622.

Wallsten, T. S. & González-Vallejo, C. C. (1994). Statement verification: A stochastic model of judgment and response. *Psychological Review, 101*, 490-504.

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: Freeman.

Yates, J.F. (1990). *Judgment and decision making*. Prentice Hall: New Jersey.

Appendix

Table A1.

Hit Rate (Relative to the High Payoff Category) for Each Experimental Condition by Observer and Block.

Observer	Condition (Figure)	Block 1	2	3	4	5
	2a	0.917	0.900	0.917	0.900	0.900
	2b	0.867	0.900	0.883	0.900	0.883
1	2c	0.917	0.867	0.917	0.883	0.950
	2d	0.917	0.883	0.917	0.933	0.900
	2e	0.850	0.850	0.850	0.833	0.850
	2a	0.900	0.917	0.883	0.917	0.900
	2b	0.883	0.883	0.900	0.900	0.917
2	2c	0.900	0.933	0.883	0.917	0.917
	2d	0.883	0.933	0.933	0.933	0.917
	2e	0.900	0.867	0.900	0.883	0.900
	2a	0.900	0.883	0.867	0.883	0.900
	2b	0.917	0.917	0.917	0.900	0.867
3	2c	0.850	0.833	0.750	0.783	0.867
	2d	0.800	0.850	0.817	0.867	0.867
	2e	0.850	0.867	0.867	0.883	0.833
	2a	0.833	0.900	0.900	0.900	0.900
	2b	0.850	0.883	0.817	0.833	0.817
4	2c	0.783	0.917	0.883	0.850	0.800
	2d	0.867	0.917	0.950	0.900	0.917
	2e	0.833	0.850	0.783	0.833	0.817
	2a	0.950	0.933	0.950	0.950	0.933
	2b	0.867	0.917	0.967	0.933	0.933
5	2c	0.950	0.967	0.983	0.950	0.933
	2d	0.983	0.933	0.950	0.917	0.933
	2e	0.883	0.933	0.967	0.950	0.933
	2a	0.900	0.867	0.917	0.850	0.867
	2b	0.883	0.883	0.833	0.883	0.867
6	2c	0.917	0.867	0.900	0.900	0.883
	2d	0.850	0.883	0.900	0.867	0.900
	2e	0.850	0.917	0.917	0.883	0.883
	2a	0.933	0.933	0.933	0.917	0.950
	2b	0.933	0.933	0.967	0.933	0.917
7	2c	0.950	0.933	0.950	0.933	0.933
	2d	0.933	0.950	0.933	0.950	0.950
	2e	0.917	0.933	0.933	0.933	0.967

(table continues)

Table A1 (continued)

Observer	Condition (Figure)	1	2	3	4	5
	2a	0.967	0.967	0.983	0.983	0.983
	2b	0.967	0.950	0.967	0.950	0.967
8	2c	0.950	0.967	0.950	0.967	0.950
	2d	0.950	0.967	0.950	0.967	0.950
	2e	0.967	0.983	0.950	0.983	0.967
	2a	0.913	0.913	0.919	0.913	0.917
	2b	0.896	0.908	0.906	0.904	0.896
Average	2c	0.902	0.910	0.902	0.898	0.904
	2d	0.898	0.915	0.919	0.917	0.917
	2e	0.881	0.900	0.896	0.898	0.894

Table A2.

False Alarm Rate (Relative to the High Payoff Category) for Each Experimental Condition by Observer and Block.

		Block				
Observer	Condition (Figure)	1	2	3	4	5
	2a	0.200	0.217	0.150	0.217	0.183
	2b	0.150	0.167	0.200	0.200	0.167
1	2c	0.217	0.183	0.200	0.183	0.200
	2d	0.233	0.233	0.167	0.167	0.217
	2e	0.117	0.117	0.150	0.100	0.133
	2a	0.133	0.150	0.217	0.233	0.167
	2b	0.167	0.200	0.167	0.133	0.167
2	2c	0.200	0.233	0.217	0.233	0.233
	2d	0.200	0.183	0.200	0.267	0.233
	2e	0.200	0.167	0.167	0.183	0.200
	2a	0.217	0.250	0.200	0.217	0.233
	2b	0.183	0.217	0.233	0.267	0.183
3	2c	0.150	0.150	0.117	0.167	0.133
	2d	0.150	0.200	0.167	0.217	0.133
	2e	0.167	0.217	0.150	0.233	0.167
	2a	0.183	0.233	0.233	0.283	0.217
	2b	0.333	0.233	0.133	0.217	0.267
4	2c	0.150	0.200	0.150	0.183	0.183
	2d	0.167	0.233	0.217	0.300	0.200
	2e	0.250	0.183	0.183	0.133	0.167
	2a	0.267	0.267	0.283	0.283	0.383
	2b	0.183	0.400	0.633	0.467	0.317
5	2c	0.350	0.450	0.383	0.433	0.367
	2d	0.350	0.267	0.300	0.233	0.200
	2e	0.150	0.333	0.333	0.317	0.333
	2a	0.267	0.183	0.283	0.250	0.167
	2b	0.167	0.183	0.217	0.200	0.167
6	2c	0.317	0.167	0.333	0.183	0.217
	2d	0.167	0.267	0.150	0.183	0.150
	2e	0.150	0.233	0.167	0.183	0.133

(table continues)

Table A2 (continued)

7	2a	0.233	0.250	0.233	0.317	0.250
	2b	0.250	0.367	0.333	0.350	0.283
	2c	0.250	0.233	0.250	0.233	0.233
	2d	0.283	0.317	0.217	0.267	0.250
	2e	0.217	0.250	0.283	0.300	0.267
8	2a	0.417	0.433	0.433	0.400	0.450
	2b	0.350	0.267	0.233	0.300	0.250
	2c	0.383	0.383	0.333	0.350	0.317
	2d	0.250	0.350	0.417	0.367	0.317
	2e	0.300	0.300	0.333	0.350	0.333
Average	2a	0.240	0.248	0.254	0.275	0.256
	2b	0.223	0.254	0.269	0.267	0.225
	2c	0.252	0.250	0.248	0.246	0.235
	2d	0.225	0.256	0.229	0.250	0.213
	2e	0.194	0.225	0.221	0.225	0.217