DISCRETIZATION OF CONTINUOUS FEATURES BY HUMAN LEARNERS

BY CORDELIA D. AITKIN

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

Jacob Feldman

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2009

ABSTRACT OF THE DISSERTATION

Discretization of Continuous Features by Human Learners

by Cordelia D. Aitkin

Dissertation Director: Jacob Feldman

Natural features are often continuous, but many models of human learning and categorization involve discrete-valued (e.g. Boolean) features. Discretization is well-known to be beneficial in machine learning, leading to faster and sometimes more accurate learning. Yet there has been little research on how human learners discretize continuous features. This dissertation investigates human discretization, focusing on two specific areas of inquiry. First is the hypothesis that discretization of a continuous parameter depends on the shape of the probability distribution underlying it, and principally on the presence of "modes" or separable peaks in the distribution. The second hypothesis is that humans create clear distinctions between discretized feature values, rather than probabilistic boundaries.

Subjects were presented with items that had feature values drawn from a mixture of Gaussian distributions, and a free sorting task was used to assess whether subjects spontaneously discretized the feature in a way that related to the underlying mixture. The relative locations of the two component Gaussians, their separation as measured by Cohen's d (the ratio of the distance between the components' means to their standard deviations), and the number of items drawn from the overall mixture were varied. Each of these factors influenced the way subjects discretized the features, while further analysis showed that the estimated mixtures were more sharply separated (higher Cohen's

d) than the original probability. This study suggests that human featural discretization involves a process akin to the estimation of mixture components in the environment, but that the separation among the components is systematically overestimated to create "cleaner" divisions than are truly present—a phenomenon that might be termed *hyperdiscretization*.

Acknowledgements

This dissertation would not have been finished without the help and support of many people.

My friends and family, in particular my parents Sue Aitkin and Rolfe Marchesano, have encouraged and cheered on all my efforts to reach this goal.

The staff of the Psychology Department and the Cognitive Area, in particular Anne Sokolowski, JoAnn Meli, and Sue Cosentino, always answered my requests for help with patience and good humor.

My outside committee member Michael Pazzani provided his expertise (of which he has much) and time (of which he has little); I am grateful for both.

My committee members Randy Gallistel and Rochel Gelman have offered encouragement and intellectual stimulation over the years. I am particularly grateful for their willingness to stay with me through the long trek from the qualifying exams until now.

My inexpressible thanks to my advisor Jacob Feldman: in addition to providing profound intellectual insights and encouragement, he has been endlessly patient and supportive.

And my deepest gratitude and love for my husband Scott, who has supported and encouraged my quest from day one.

Dedication

This thesis is dedicated to my daughter, Alexandra.

Table of Contents

List of Figures

Chapter 1

Introduction

Natural physical parameters, such as distance, time, and mass, are often treated as intrinsically continuous, as are primitive psychophysical parameters like orientation and size. But many models of human cognition assume discrete-valued (e.g. binary) features, with qualitative separation between distinct levels or values (Shepard, Hovland, & Jenkins, 1961; E. E. Smith, Shoben, & Rips, 1974; Tversky, 1977). Additionally, although several models allow for continuous-valued features, the support for such models generally uses discrete values (e.g. Medin & Schaffer, 1978; Nosofsky, 1986). In other models, these distinct levels often become the primitive elements of more complex compositional systems, and thus form a key component of the symbolic processing that underlies much of cognition (e.g. Anderson, 1991). Yet exactly how the values of a continuous feature (e.g. *size*) are aggregated or binned to yield a discretized feature (e.g. *big/small*)—the process of *discretization*—has received little attention in the psychological literature. This thesis investigates some very basic questions about how this process works.

Because it discards distinctions among points within a discrete class, discretization inherently involves a loss of information. However, several classic results indicate that people are not able to use all the information available. For example, Garner and Hake (1951) found that no matter how discriminable their stimuli, subjects were unable to learn names for more than a small number of discrete values of an acoustic signal. More broadly, Miller (1956) famously proposed that human cognizers face severe limits on the number of distinct states of any variable or process they can entertain (e.g., 7 ± 2). As it ultimately aims to *minimize* information loss, discretization seems like a natural heuristic for human cognition.

1.1 Theoretical Motivation

There are many different ways to discretize a continuous feature. A feature can simply be split into equal intervals (a common choice in the applied literature), or it can be divided so that each bin has equal probability mass (as in a median split, which divides an interval into two bins at the 50th percentile). The latter idea is very simple, but contains the grain of a more general idea I wish to develop: that how a discretization is chosen might depend on the way the continuous values are distributed in the examples, or in the environment from which they are drawn.

If a given continuous variable is distributed uniformly in the environment, there is little basis for selecting dividing points (called *cutpoints*), so any discretization would be somewhat arbitrary. But if the variable's underlying probability density function is "spiky" or conspicuously multimodal, it seems desirable for the cutpoints to reflect the natural divisions between the modes (Feldman, 2009), which would suggest discretization may subserve efficient coding. Indeed, a robust area of machine learning research has found discretizers that use the statistical information in the dataset produce the best combination of speed and accuracy.

In other areas of research, evidence indicates natural parameters are distributed non-uniformly, or unevenly; as such, human discretization would be most efficient if it captured rather than obscured that unevenness. Finally, two broad areas of psychological research indicate people may be able to use statistical information in the environment. First, ample research has shown that low-level processes in vision and audition are uniquely suited to take advantage of these uneven statistics. Second, research in areas such as categorization indicate people may be sensitive to statistical information in higher-order processes as well. Thus, human discretization may also be able to use the statistical information about a feature in order to create a relatively accurate symbolic model of the world. In order to make this point clearer, I will now review these areas of research.

1.1.1 Machine Learning

Discretization has been much studied in machine learning, where aggregation of "continuous" features (in practice meaning features with a large number of used values) into discrete features (with a small number of distinct values) is well known to yield both a reduction in computational complexity as well as, in many cases, an improvement in performance (Dougherty, Kohavi, & Sahami, 1995; Grabczewski, 2004). Like computers, humans have limited computational resources, and so human discretization might well follow principles similar to those advanced in the computational literature.

As alluded to above, the simplest discretizers take the feature and divide it into equal sections based on either range (equal width) or the number of occurrences (equal frequency). However, these methods may obscure useful information like category clusters (Dougherty et al., 1995; Kurgan & Cios, 2004; Yang & Webb, 2002). Therefore, researchers have developed discretizers which reduce the loss of important or helpful data during the discretization process. Various statistical and information-theoretic tests are used to determine if a cutpoint is useful or not; some of the methods include minimum description length (Fayyad & Irani, 1993; Friedman & Goldszmidt, 1996), entropy (Kohavi & Sahami, 1996), and mutual information (Kurgan & Cios, 2004).

Although early research focused on discretizing one feature at a time, more recent work has explored the value of discretizing a number of features simultaneously, as focusing on one feature at a time may mask interesting information in the data (Bay, 2001; Ludl & Widmer, 2000; Monti & Cooper, 1999; Wang & Liu, 1998). Bay (2001) uses the example of XOR data, where items in both categories are spread approximately evenly across any one feature. However, if one examines the data using both features simultaneously, the four distinct clusters become obvious.

1.1.2 Natural Feature Distributions

Many theorists have speculated that cognitive representations are in some way based on correlational or regular structure in the world (Anderson, 1991; Rosch, 1978). In a similar vein, I speculate that discretization is sensible when it reflects the "natural modes"

(Richards & Bobick, 1988) in the environment, specifically the distinct components (if any) in the distribution of values of a continuous feature. In this sense, my motivation is similar to research in natural image statistics. Here, the visual system has been found to reflect the statistics of the natural world in a variety of senses and domains (Alvarez & Oliva, 2009; Field, 1989; Simoncelli & Olshausen, 2001). Thus, humans may be sensitive to the uneven, multimodal nature of the statistics of the continuous world when discretizing as well.

1.1.3 Higher-Order Use of Uneven Distributions

Existing evidence of how humans use uneven distributions in categorization and other higher-order processes is rather indirect and contentious. The first set of evidence comes from developmental psychology. Kemler Nelson and collegues have used the free-sort task to investigate how children categorize. (Free-sorting involves giving the subjects items and asking them to group the items that "go together." "Go together" is deliberately not defined by the experimenters, and subjects are not given feedback as to whether the groups are "correct" or not.) They found that children tend to create "family resemblance" (FR) categories; that is, they tend to group items that are globally similar in that the items have co-occuring feature values, although not all co-occuring values coincide in every item (Kemler & Smith, 1979; L. Smith, 1989). This contrasts with adults, who tend to sort items by a single feature even if that feature divides an FR category (Ahn & Medin, 1992). However, Berger and Hatwell (1996), using free-sorting of haptic rather than visual cues, found that the developmental difference may be in the level of processing; that is, adults are more likely to analyze stimuli using high-level information, while children are more likely to use low-level information. This is supported by additional work of Kemler Nelson and colleagues (Foard & Kemler Nelson, 1984; J. Smith & Kemler Nelson, 1984), who found that under speeded conditions, holistic processing of certain types of visual stimuli was more likely than analytic processing in adults.

However, when more feature values co-occur and so long as there is no one universal feature, adults are more likely to create FR categories that any other type of category

(Ahn & Medin, 1992; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Ward, Foley, & Cole, 1986). Additional research has shown that people can use co-occurring feature values in other categorization and learning tasks (Billman & Knutson, 1996; Crawford, Huttenlocher, & Hedges, 2006). As an example, Clapper and Bower (1994) used line drawings of insects and asked subjects to list features that distinguished a particular insect from others like it; subjects quickly stopped listing co-occuring features. Thus, research in higher-order use of statistical information suggests subjects may be able to notice clusters of values within a single feature, and therefore produce a discretization of the feature which reflect those clusters.

1.2 The Form of the Discretization

There are two possibilities for the type of cutpoints people might use: the cutpoints may be either "fuzzy" or "clean." Research in other areas of psychology suggest the former. The classic example is Weber's Law, which states that the discriminability of two values depends on the value. More recently, timing (Malapani & Fairhurst, 2002) and nonverbal number representation (Cordes, Gelman, Gallistel, & Whalen, 2001; Whalen, Gallistel, & Gelman, 1999) have been shown to have a scalar variability. However, the results from machine learning show that an informative discretization which can increase accuracy and reduce effort depends in part on the discretization being exact (Grabczewski, 2004). Thus, in order to understand how much help human discretization offers in processing and modeling the environment, it is necessary to determine how exact human discretization is.

1.3 Summary of the Motivation

There is a good deal of information available in the natural world, some of which is necessary in order to interact successfully with the world and some of which is not. In order to efficiently interact with the world, humans must have ways to efficiently process that information. For features in particular, one way to make them more efficient is to reduce the number of values without catastrophically reducing the amount of information

contained in the features: discretization. Although well-studied in machine learning, the process by which humans transform a continuous feature into a discrete feature has not been examined. In a first step towards understanding how continuous features become discrete, symbolic values in the stream of cognition, this thesis examines some of the basic principles of human discretization.

Chapter 2

Experimental Approach

Clearly, an unexplored problem will have many possible avenues of approach. This chapter lays out the reasoning behind the particular experimental procedure which was chosen.

2.1 Discretization and Environmental Statistics

Discretization might be seen as a type of (unsupervised) categorization or classification, if one regards the many potentially distinguishable levels of the underlying continuous variable as items to be categorized. In this case the discrete levels would simply be learned classes as in any other categorization problem, and the process of discretization would simply be a special case of the well-studied problem of category formation. I view discretization as a distinct and indeed more basic process, contributing as it does to the formation of the features themselves, and thus establishing the basic perceptual/cognitive vocabulary over which later representations are built. This is complementary to the process by which psychological features are chosen (or created), as explained by the fundamental work Schyns & colleagues (Schyns, Goldstone, & Thibaut, 1998; Schyns & Rodet, 1997). Schyns & collegues have found that the process of categorization can influence which aspects of an object (e.g. segments of a boundary) subjects use as features. The features Schyns colleagues tend to be binary (there/not-there); additionally, the full process is still poorly understood. I view feature discretization as a complementary process, in which the basic (continuous) perceptual or psychophysical features are transformed to become the discrete, symbolic features that are often assumed to be employed by later cognitive processes.

If human discretization does attempt to minimize information loss, discrete values of

an uneven distribution should correspond to the modes in that distribution. To explore this idea, objects with features whose values were distributed according to *mixtures* of multiple unimodal density functions were presented to subjects for sorting. A mixture distribution or mixture model is the normalized (i.e, weighted in such a way that the integral remains 1) sum of several component distributions, sometimes called sources (McLachlan & Basford, 1988). For example, the probability density function $p(x)$ of the continuous parameter x might be the mixture of K components $g_1(x) \ldots g_K(x)$,

$$p(x) = \sum_{i=1}^{K} w_i g_i(x), \tag{2.1}$$

in which w_i are the mixing proportions, and the i-th component source has a distinct mean μ_i and standard deviation σ_i.

One question to ask about human discretization is whether discretization is generally easier when the components are more separable. A natural measure of separation among components is the ratio of the distance between their means to their spread, sometimes called Cohen's d

$$d = \frac{|\mu_1 - \mu_2|}{\sqrt{(\sigma_1^2 + \sigma_2^2)/2}}, \tag{2.2}$$

commonly used in the statistical literature as a measure of effect size (Cohen, 1988). Cohen's d is high when the distributions are well-separated relative to their spread, and is low when they overlap substantially, in which case the mixture may not be visibly multimodal; Fig. 2.1 shows two examples of different levels of Cohen's d.

2.2 General Motivation

The overall goal of the experiments was to understand the basic mechanisms of discretization. More specifically, these experiments tested the hypothesis that discretization of a single feature is sensitive to the underlying environmental distributional structure and involves an attempt to recover the mixture components. There are two major aspects to this question:

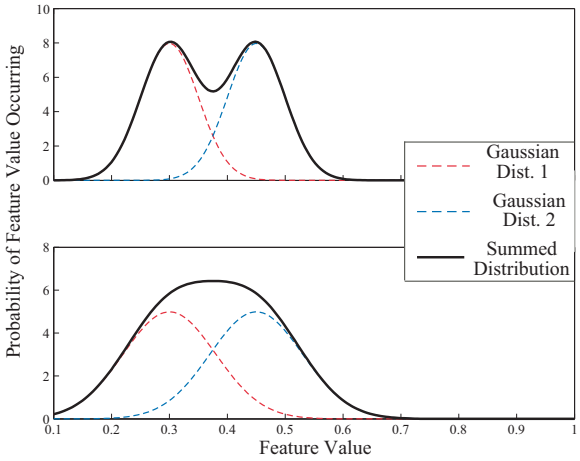

Figure 2.1: Mixtures of two Gaussian distributions. The top panel shows moderate separation (Cohen's $d = 3$), while the lower panel shows low separation (Cohen's $d = 1.875$).

1. will subjects' cutpoints be hard cutpoints (as in machine learning), or will they be more like the probabilistic estimations of the generating distribution?

2. will subjects be influenced by the distribution such that the location of their cutpoint matches the cutpoint of the distribution?

The first question concerns how clean the subjects' cutpoints are. To make "clean" clearer, consider an example. Assume we have a continuous variable x, the values of which $\{x_1, x_2, x_3, ...\}$ occur according to some generating distribution. We would like to map x onto a discrete variable $\overline{x}$, such that this related discrete variable has two values $\overline{x}_1$ and $\overline{x}_2$. The division between $\overline{x}_1$ and $\overline{x}_2$ occurs at value x_i of the original feature x. One source of fuzziness in the cutpoint comes from the selection of x_i; if x_i is chosen by a stochastic process, the cutpoint between $\overline{x}_1$ and $\overline{x}_2$ will be fuzzy. However, even if the process is not stochastic, the cutpoint may still be fuzzy. Consider the value $x_i - \epsilon$ of the original x such that $x_i - \epsilon$ is very close to x_i while still perceptually discriminable from x_i. When assigning $x_i - \epsilon$ to a discretization value, there are two possibilities: either this value will always be $\overline{x}_1$ (a clean discretization); or it will be $\overline{x}_1$ with some probability p and $\overline{x}_2$ with probability $1 - p$ (a fuzzy discretization). While the latter

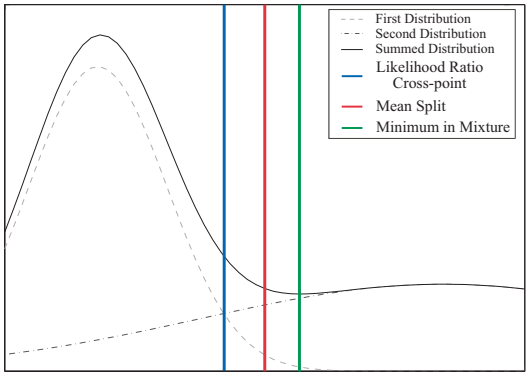

Figure 2.2: Example of deconfounded environmental cutpoints. Blue is the point of equal likelihoods of the two Gaussians; Red is the mean split; Green is the minimum in the mixture density

possibility is the form other human cognitive behaviors generally take (e.g. Weber's Law), the former is the discretization schema used in machine learning.

The second question can be rephrased as: do subjects recover the components of a mixture? Estimation of the sources in a mixture in the general case is a complex process (McLachlan & Basford, 1988), but in this situation several simple principles suggest themselves. Subjects might place cutpoints exactly between the component means (a *mean split*), place them so equal numbers of stimuli are on both sides (a *median split*); place them at a minimum in the mixture density; or place them at the point that is equiprobable under both source distributions (the *likelihood ratio cross-point*), which is the Bayes optimal solution (assuming equal priors, as in these experiments) (see Fig. 2.2). When standard deviations are equal (homoscedastic case, Exp. 1A and Exp. 2), all these hypothetical cutpoints coincide; the goal of those experiments is thus more fundamental, to establish that discretization is sensitive to the structure of the generating distribution. When the standard deviations are unequal (heteroscedastic case, Exp. 1B) these potential cutpoints are deconfounded, allowing a more fine-grained analysis.

2.3 General Experimental Procedure

Subjects were asked to freely sort objects whose feature values were generated from a mixture distribution defined over one or two salient (quasi-)continuous parameters. Two features were selected under the following restrictions: salient, unidimensional (unlike e.g. color), and naturally bounded at both ends (unlike e.g. length). The features selected were *luminance* (perceived reflectance, ie lightness) and *aspect ratio*, the ratio of a shape's shorter dimension to its longer, a simple shape parameter known to be psychologically meaningful (Feldman & Richards, 1998). Given these two features, an ellipse was chosen to be the sorted object. Aspect ratio ranges in principle from 0 (a line segment) to 1 (in ellipses, a circle); however, to avoid the degenerate case of a line segment, aspect ratios varied only from 0.1 to 1 in the experiments.

For all experiments, stimulus sets were drawn from mixtures of two Gaussians defined over the chosen feature(s). The two source components had equal weights (w_i) and standard deviations that were either equal (Exp. 1A and Exp. 2, homoscedastic case) or unequal (Exp. 1B, heteroscedastic case). One of the main questions of interest was whether the separation of the two components would influence the way subjects discretized, which would suggest a rational data-driven discretization process involving some estimation of the source components. Hence the source components were chosen in such a way as to vary Cohen's d over a wide range, from low values entailing heavily overlapping distributions to high values entailing well-separated ones.

A second question of interest concerns whether humans are optimally influenced by the data. As shown in Fig. 2.2, there are at least three possible cutpoints in the Gaussian mixture. However, these distributional cutpoints are fuzzy cutpoints; thus, if humans are optimally recovering the distribution, the cutpoints will be as well.

Two other basic parameters of the stimulus set were manipulated in the experiments. First, the location of the two source components (i.e., their means μ_i) in the parameter space was also manipulated, as features are not always equally discriminable across the space of values (e.g. loudness, weight); although there is no clear increasing or decreasing direction in aspect ratio or luminance, there is certainly a potential for asymmetry, and

thus for discretization cutpoints to "drift" in one direction or another, suggesting some kind of prior bias. Second, as a way of manipulating the quantity of data available to the subjects when making their discretization choices, the number n of stimuli presented in a block was also varied.

2.3.1 General Analysis

This thesis concerns three substantially independent aspects of subjects' discretization procedures, which were quantified separately as dependent measures. First, it was determined where the subjects placed the cutpoints by identifying the cutpoint that best classified the subjects' own responses. Second, the *discretization error* (**DE**) measures how cleanly or "discretely" subjects actually discretized, that is, how strongly the ensemble of observed values induced them to form a clear boundary and apply it. To quantify this, the stimulus items that were still misclassified relative to this cutpoint determined above (i.e. the number of ellipses on the "wrong side" of the subject's own apparent boundary) were counted; DE = 0% would mean a perfectly clean boundary. Finally, *ideal error* (**IE**) measures the subject's "objective" error in separating the source distributions. This can be thought of as a measure of the subjects' performance relative to an ideal Bayesian observer who knew the form of the source components in advance but not their locations. IE = 0% means an exact boundary that matches the likelihood ratio cross-point of the underlying distribution. Thus the IE measures success in determining the source components, while the DE simply measures the strength of discretization without reference to the true sources.

2.4 Summary of the Experimental Approach

This thesis addresses two of the basic questions about human discretization: do people cleanly discretize, and does the distribution of continuous feature values affect where people separate discrete values. To answer these questions, the general form of the experiments involved free-sorting. Two features were used: aspect ratio and luminance. Three parameters of the generating mixture model were varied: separation (as measured

by Cohen's d), location, and number of items drawn from the mixture. Three dependent measures assessing different aspects of the discretization were analyzed: the location of the subject's cutpoint, the number of errors relative to that cutpoint (DE), and the number of errors relative to the likelihood ratio cross-point in the mixture (IE). The next chapter presents the details of the first experiment.

Chapter 3

Experiment 1 - One Feature

The first experiment involved discretizing a feature while all other features remained at a constant value. Stimuli were ellipses of either various aspect ratios or various percentages of grey (with 0%=black and 100%=white) as drawn from mixtures of two Gaussians. Experiment 1A tested the idea that the separation in the underlying distribution, as measured by Cohen's d, affected discretization accuracy as measured by Discretization Error and Ideal Error (described in Section 2.3.1). Thus, the crucial measure manipulated in Exp. 1A was the separation (Cohen's d). Experiment 1B looked more closely at what particular aspect of the distribution affected discretization accuracy; thus, the additional measure manipulated was differentiation of various environmental cutpoints, by using heteroscedastic distributions. In Experiment 1, three parameters were modified: separation (as determined by Cohen's d); location in feature space (as determined by the average of the two means); and number of items in a block. The statistical parameters were the same for both aspect ratio and luminance, and so will be described fully now.

3.1 Stimuli and Procedure

3.1.1 Stimulus Parameters

Separation

Three levels of separation were tested in Exp. 1A. The two means in each mixture were held at a constant distance from each other, so change in Cohen's d was determined entirely by changing the standard deviation. In Exp. 1A, one distance between means and three (shared) standard deviations were chosen.

Six levels of separation were tested in Exp. 1B. In Exp. 1B, the distance between the two means was not held constant. Thus, Cohen's d depended on the two different standard deviations and on the distance between the means. In Exp. 1B, three distances and two sets of standard deviations were chosen.

A list of the means, standard deviations, and separation as measured by Cohen's d for both Exp. 1A (homoscedastic) and Exp. 1B (heteroscedastic) can be found in Appendix A.

Location

Both Exp. 1A and Exp. 1B had six locations, as defined by the average of the two means. As the means in Exp. 1A were always the same distance apart, the six locations involved moving both means. In Exp. 1B, however, three locations involved moving only the larger mean, and the other three locations involved moving only the smaller mean.

Number of Items

With more items drawn from the mixture, subjects would have more data about the mixture. Therefore, each mixture was presented three times, with a different number of items each time. Exp. 1A presented either 10, 20, or 40 stimuli, while Exp. 1B presented either 20, 35, or 45 stimuli.

3.1.2 Overall Design

These three factors (separation; location in feature space; number of items) were completely crossed within subjects. For Exp. 1A (homoscedastic), this yielded a total of 54 blocks. For Exp. 1B (heteroscedastic), as location and separation are partially confounded, a complete crossing yielded 36 blocks. All blocks were presented in random order. The experimental session lasted approximately one hour.

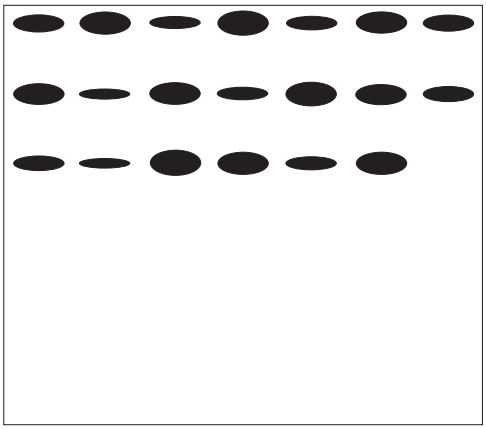

Figure 3.1: Example of a "group" screen in a homoscedastic aspect ratio sort, Exp. 1

3.1.3 General Procedure

The procedure was the same for all the experiments. For each block, the subject studied a group of ellipses on the computer screen, arranged in a grid in random order, and mentally sorted them into two groups ("narrower"/"rounder" or "lighter"/"darker"; see Fig 3.1) . When the subject felt comfortable about how s/he would assign the ellipses to the two groups, the subject pressed any key on the keyboard. Each ellipse was then presented individually in (a new) random order. The individual ellipses were larger than the ellipses on the group screen (where they had been uniformly reduced in size to fit on one screen). The subject was asked to press one key if s/he had decided the ellipse was in one group, and another key if s/he had decided the ellipse was in the other group. After each response, a blank screen appeared and the subject would press the space bar to start the next trial. After the subject classified all the ellipses from that group, a new block with a new set of ellipses drawn from a new mixture would begin.

Before the start of the experiment, subjects were given written instructions explaining the procedure, including that sorting should be based only on the ellipses that were visible. After the subject had read the instructions but before starting the experiment itself, the experimenter re-emphasized that the subject would be seeing a broad range of ellipses over the course of the experiment, but that for any given group, they should

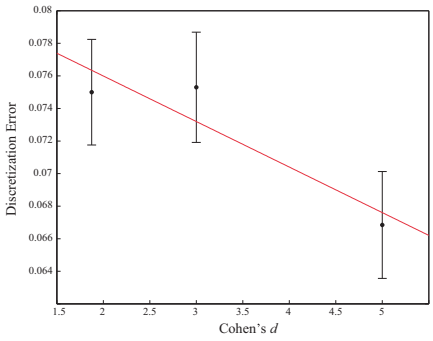

Figure 3.2: Discretization error (DE) as a function of separation (Cohen's d), Exp. 1A.1. Error bars here and in other figures are ± 1 s.e.

decide the sorting based only on the other ellipses in the current group.

3.2 1A: Homoscedastic Modes

3.2.1 1A.1: Aspect Ratio

Subjects and Design

Subjects were 23 undergraduate students receiving class credit in return for participation, and were naive to the purpose of the experiment. Stimuli were ellipses of various aspect ratio with equal standard deviations as described above; luminance was held constant (0%, black).

Results

Mean Ideal Error over all subjects and all blocks was 17.4%; mean Discretization Error over all subjects and blocks was 7.1%.

There was a significant effect of Cohen's d on both DE ($F(1, 1240) = 11.3, p < 0.001$; Fig. 3.2) and IE ($F(1, 1240) = 30.0, p < 0.0001$; Fig. 3.3)[1], such that higher values of Cohen's d were associated with lower error rates.

[1] As the independent variables are numeric rather than categorical, unless otherwise indicated, I used linear regression to test their effects here and elsewhere.

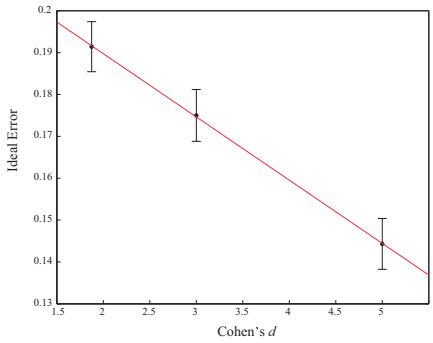

Figure 3.3: Ideal error (IE) as a function of Cohen's d, Exp. 1A.1.

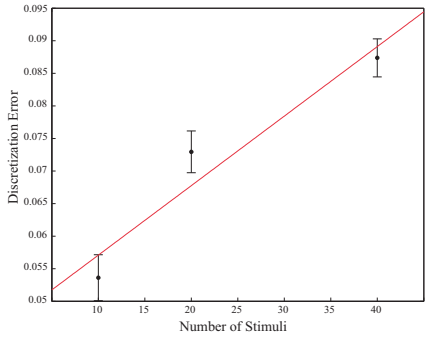

Figure 3.4: DE as a function of number of stimuli, Exp. 1A.1.

Surprisingly, there was no significant effect of n on IE; however DE increased with n ($F(1, 1240) = 27.7, p < 0.0001$; Fig. 3.4).

To compare the subjects' cutpoints to ideal cutpoints, for each block the distance and direction from the subject's cutpoint to the ideal was calculated. As can be seen in Figure 3.5, the distance subjects moved the cutpoint depended on both the location of the underlying distribution and the direction the cutpoint was moved (2-way ANOVA. Location: $F(5, 266) = 2.7, p < 0.03$; direction: $F(1, 266) = 5.6, p < 0.02$, interaction: $F(5, 266) = 15.7, p \ll 0.00001$). That is, for distributions of feature values centered closer to 1, subjects tended to move cutpoints downwards relative to the Bayes-rational cutpoint, and for distributions centered closer to 0, subjects tended to move cutpoints upwards relative to the Bayes-rational cutpoint. Additionally, the further the mixture

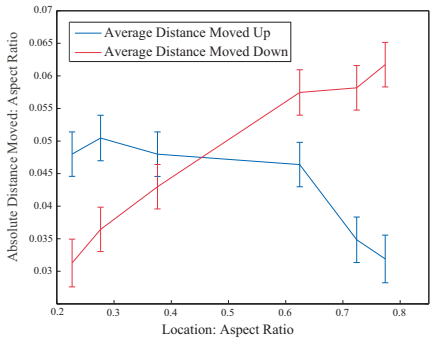

Figure 3.5: Change in cutpoint as a function of location in feature space, Exp. 1A.1.

was from the overall mean of the feature space, the further the subjects moved the cutpoint.

3.2.2 1A.2: Luminance

Exp. 1A.2 is a replication of Exp. 1A.1 using a different feature (luminance), to confirm the results were not specific to aspect ratio.

Subjects and Design

Subjects were 22 (new) undergraduate students receiving class credit in return for participation, and were naive to the purpose of the experiment. One subject was dropped due to a recording error, and one subject was dropped because both IE and DE were more than three standard deviations from the mean error rates: a total of 20 subjects were analyzed. Stimuli were ellipses of constant aspect ratio (0.5) and varying luminance drawn from homoscedastic distributions as described in Section 3.1.

Results

Mean overall IE was 16.8% and mean overall DE was 7.1%. Cohen's d had a significant effect on DE ($F(1, 1078) = 9.8, p < 0.005$; Fig. 3.6) and on IE ($F(1, 1078) = 27.2, p < 0.0001$; Fig. 3.7), such that higher values of Cohen's d led to lower error rates. The number n stimuli did not have a significant effect on IE, but did significantly increase

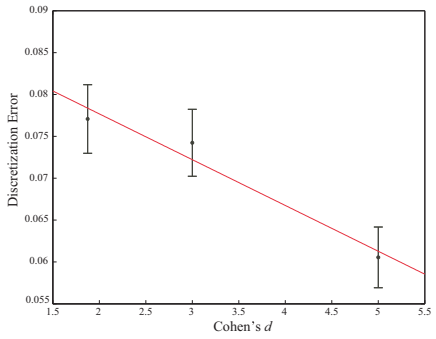

Figure 3.6: Discretization Error as a function of Cohen's d, Exp. 1A.2.

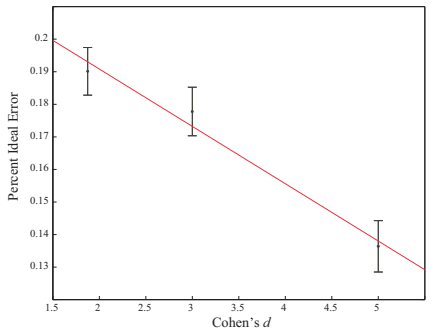

Figure 3.7: Ideal Error as a function of Cohen's d, Exp. 1A.2.

DE ($F(1, 1078) = 39.8, p < 0.0001$; Fig. 3.8)

For cutpoint movement, the distance an average cutpoint moved depended on location of the underlying distribution, and whether the cutpoint was being moved up (toward white) or down (toward black), such that cutpoints of distributions further from the mean of the overall feature space were moved further than cutpoints of distributions closer to the mean of the overall feature space, and cutpoints at all locations were moved toward the mean of the entire feature space (2-way ANOVA: interaction between distance and direction: $F(5, 217) = 5.64, p < 0.005$; Fig 3.9).

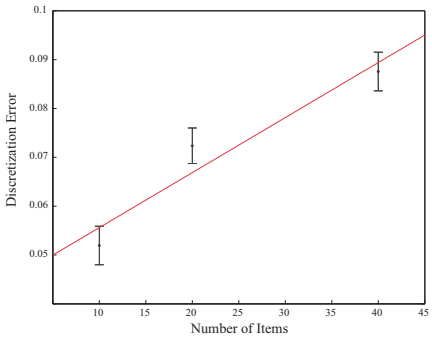

Figure 3.8: DE as a function of number of stimuli, Exp. 1A.2.

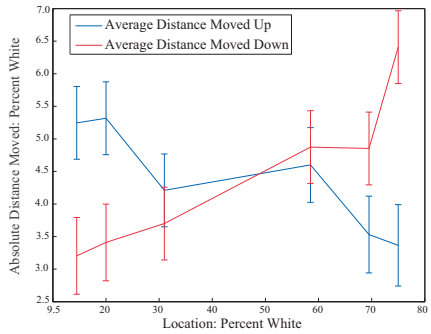

Figure 3.9: Change in cutpoint as a function of location in feature space, Exp. 1A.2.

3.2.3 Analysis: Homoscedastic Modes

These findings, in particular the significant influence of the separation between modes (Cohen's d), suggests that subjects are influenced by the underlying distribution. Although the number n of stimuli provided more data about the underlying distribution, this additional data did not allow subjects to recover the underlying cutpoints more accurately (IE); however, subjects did create fuzzier distributions (DE) with more data, possibly reflecting the shape (if not the location) of the underlying mixture more accurately. The cutpoint movement suggested that subjects have a bias towards the mean of the overall feature space, despite explicit instructions to focus only on the current feature values. As the overall results suggested that subjects are influenced by the distribution of current values, the next set of experiments looked at what aspect of the

underlying distribution may have the most influence.

3.3 1B: Heteroscedastic Modes

Exp. 1A showed that subjects' discretizations were somewhat influenced by environmental distributions of feature values: more separation in the distributions produced better results, both in the subjects' discretization and compared to the original distribution. However, it did not indicate which environmental cutpoint subjects may be aiming to replicate; as mentioned in Section 2.2, there are several possible environmental cutpoints that are available, including the mean split and the minimum in the mixed density. Exp. 1B tested how subjects would discretize when the standard deviations of the two source components were unequal, which deconfounds several candidate cutpoints: the midpoint between the two means of the components, the likelihood ratio crosspoint, and the minimum in the mixed density.

3.3.1 1B.1: Aspect Ratio

Experiment 1B.1 involved heteroscedastic modes and asked the subjects to sort by aspect ratio.

Subjects and Design

Subjects were 22 (new) naive undergraduate students receiving credit in return for participation. The stimuli were ellipses of constant luminance (0%, black) and varying aspect ratio, drawn from heteroscedastic mixtures described in Section 3.1.

Results

Mean overall Ideal Error was 15.3%; mean overall Discretization Error was 4.9%. Cohen's d had a significant effect on both DE ($F(1,790) = 11.2, p < 0.001$; Fig. 3.10) and IE ($F(1,790) = 77.4; p \ll 0.0001$; Fig. 3.11), such that higher levels of Cohen's d led to lower error rates. The number of stimuli did not have a significant effect on IE, but DE was significantly higher with more items (DE: $F(1,790) = 12.4; p < 0.001$, Fig. 3.12).

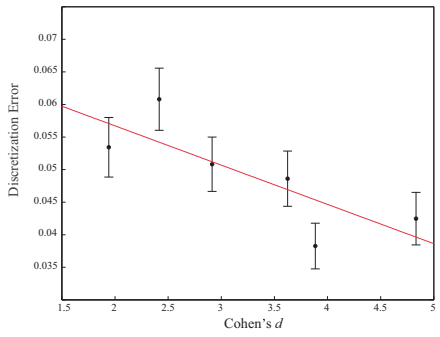

Figure 3.10: Discretization Error as a function of Cohen's d, Exp. 1B.1.

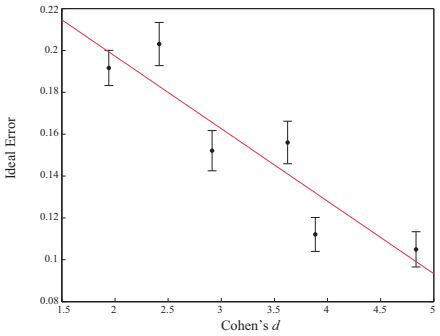

Figure 3.11: Ideal Error as a function of Cohen's d, Exp. 1B.1.

Comparing subjects' cutpoints to the underlying equal-likelihood cutpoints showed that subjects tended to move their cutpoints toward the mean of the entire feature range, and that the further the original distribution was from the mean of the feature range, the further the average cutpoint was moved (2-way ANOVA: interaction of location in feature space and direction of movement: $F(11, 504) = 14.9, p \ll 0.0001$; Fig 3.13).

The primary goal of this set of experiments was to try to understand how subjects chose their cutpoints. As such, subjects' cutpoints for each block were compared to three statistical cutpoints: the midpoint between the two means, the likelihood ratio crosspoint, and the minimum in the mixed density. The absolute distance (i.e. movement regardless of direction) from the subject's cutpoint to the environmental cutpoint

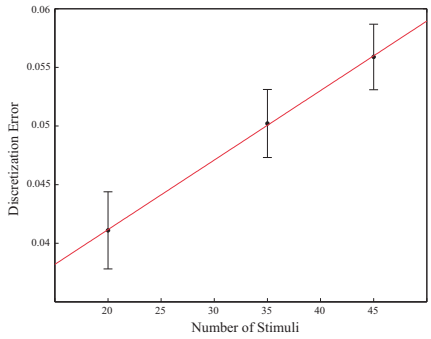

Figure 3.12: DE as a function of number of stimuli, Exp. 1B.1.

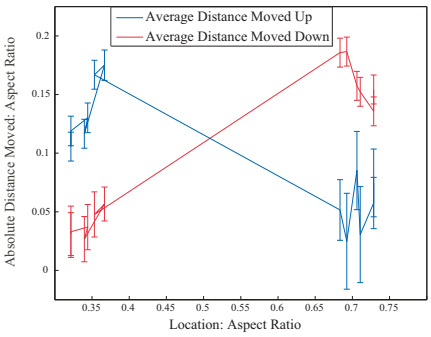

Figure 3.13: Change in cutpoint as a function of location in feature space, Exp. 1B.1.

was calculated for each block. Subjects' cutpoints were closest to the average of the two means (ANOVA, $F(2, 2373) = 34.7, p \ll 0.00001$), suggesting a simple discretization heuristic; Figure 3.14 shows the average absolute distance between subjects' cutpoints and each of the three possible environmental cutpoints.

3.3.2 1B.2: Luminance

Experiment 1B.2 involved heteroscedastic modes and asked subjects to sort by luminance.

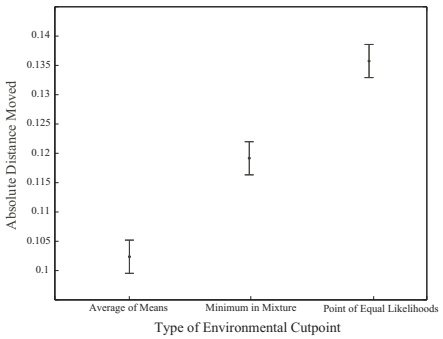

Figure 3.14: Average distance between subject's cutpoint and environmental cutpoint, by type of environmental cutpoint, Exp. 1B.1.

Subjects and Design

Subjects were 23 (new) undergraduate students receiving credit in return for participation, and were naive to the purpose of the experiment. One subject had error rates that exceeded three standard deviations from the mean error rates: a total of 22 subjects were analyzed. Stimuli were ellipses of constant aspect ratio (0.5) and varying luminance drawn from heteroscedastic distributions as described in Section 3.1.

Results

Mean overall IE was 12.0%, while mean overall DE was 4.2%. Cohen's d had a significant effect on DE ($F(1, 790) = 11.7, p < 0.001$; Fig. 3.15) and on IE ($F(1, 790) = 17.5, p < 0.0001$; Fig. 3.16), such that higher levels of Cohen's d led to lower error rates. The number of stimuli did not have a significant effect on IE, but DE was significantly higher with more items ($F(1, 790) = 12.0, p < 0.001$; Fig. 3.17).

Comparing subjects' cutpoints to the underlying equal-likelihood cutpoints showed that subjects tended to move their cutpoints toward the mean of the entire feature range, and that the further the original distribution was from the mean of the feature range, the further the average cutpoint was moved (2-way ANOVA: interaction of location in feature space and direction of movement: $F(9, 363) = 29.45, p \ll 0.001$; Fig 3.18).

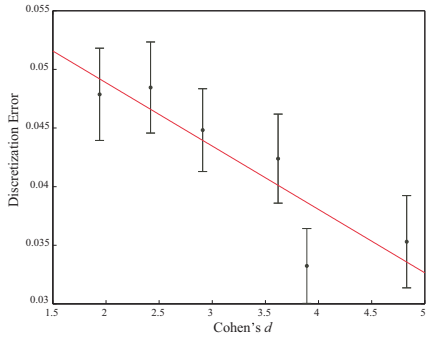

Figure 3.15: Discretization Error as a function of Cohen's d, Exp. 1B.2.

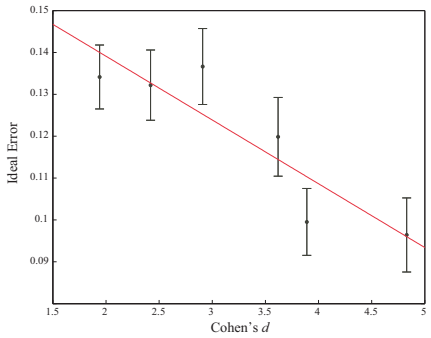

Figure 3.16: Ideal Error as a function of Cohen's d, Exp. 1B.2.

Comparing subjects cutpoints to the three possible environmental cutpoints replicated the results from Exp. 1B.1: subjects' cutpoints were closest to the average of the two means (ANOVA, $F(2, 2373) = 13.7, p \ll 0.001$; Fig 3.19).

3.4 General Discussion, Exp. 1

As discussed in Chapter 1, these experiments were motivated by the need to answer some basic questions about how humans create discrete features from continuous ones. The first experiment explored the idea that humans may discretize a feature according to the various frequencies of the feature's values in the environment. As noted previously, if an environment is sufficiently non-uniform, a discretization that captures that non-uniformity can increase the accuracy of interactions with that environment (Alvarez &

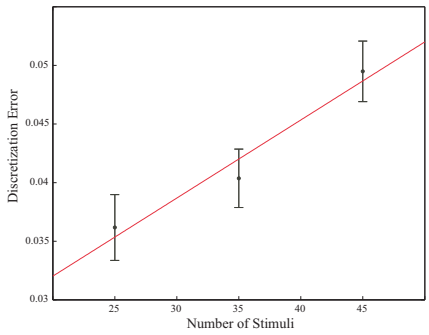

Figure 3.17: DE as a function of number of stimuli, Exp. 1B.2.

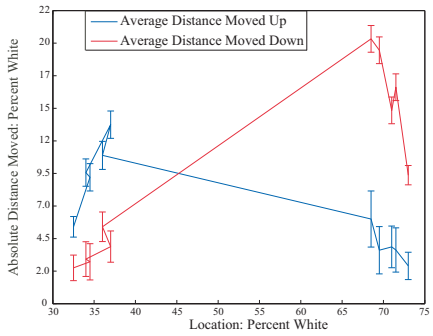

Figure 3.18: Change in cutpoint as a function of location in feature space, Exp. 1B.2.

Oliva, 2009; Monti & Cooper, 1999). The main hypothesis of this dissertation is that the formation of discrete levels of a continuous feature is related to the estimation of mixture components, in that each resulting symbol (i.e., discrete level) is intended to correspond to one distinct source of observations. The results of Exp. 1 demonstrated that subjects' discretizations are sensitive to the statistics of the environment; subjects do not divide the continuous feature evenly, nor arbitrarily, but in a manner that reflects the way it is distributed in their environment. In all of Experiment 1, subjects gave responses that reflected the underlying distribution rather than an arbitrary method of sorting. Additionally, their discretizations were both cleaner and more accurate when the underlying sources were more distinctly separated (as parameterized by Cohen's d).

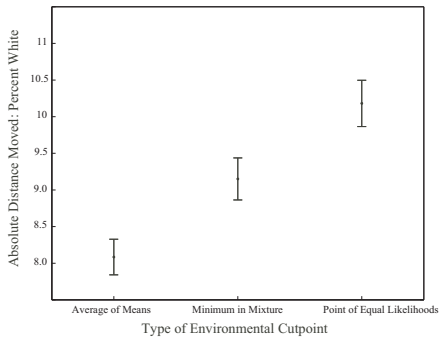

Figure 3.19: Average distance between subject's cutpoint and environmental cutpoint, by type of environmental cutpoint, Exp. 1B.2

As more data would generally be expected to produce higher accuracy, it was unusual to find that an increase in the number of items did not produce a more accurate discretization. However, subjects did produce a fuzzier cutpoint with more items. It is not quite clear from the current results why these results occurred. There are several possible explanations. For example it may be that (in the spirit of the motivation for this work), subjects were not able to use all the available information. That is, remembering and responding to 40 items (such as aspect ratios or luminance levels) is more difficult and takes more processing power than remembering and responding to 10. Alternatively, it may be that people were using all the available information; more data provided more information about the fuzziness of the underlying distribution. However, determining the cause of this effect cannot be clarified by these experiments and must await further research.

The next question was: do subjects discretize in an ideal manner? An ideal manner would accurately reflect all the information available in the environment. For example, if the subjects were able to retrieve the underlying component distributions, a natural cutpoint would be at the point of equal likelihoods. Alternatively, if the subjects were able to retrieve the overall distribution (but not the individual components), an ideal cutpoint would be at the dip between the two means. However, several results indicated that subjects, although influenced by the environmental information, were not ideal. First, the Discretization Error was generally lower than the Ideal Error, suggesting that

subjects were able to make fairly clean distinctions between groups even when those groups did not exactly match the underlying distribution. Second, the movement of the cutpoints indicated a mean-drift; the average subjective cutpoint tended to be between the mean of the full range of values and the environmental cutpoint, suggesting subjects were influenced by an awareness of the full range of values even when there was only a sub-range of values visable. Finally, the results from Exp. 1B indicate that subjects appear to be discretizing at the average of the two means, rather than either of the cutpoints described above.

As explained in Section 2.2, there are two primary questions motivating this work: are people influenced by statistical information in their environment when determining discrete values of continuous features, and do people create hard boundaries between the discrete values? As shown by these four experiments, the level of separation, as represented by Cohen's d, influences the determination of discrete values; the higher the separation in an underlying distribution, the more likely people are to put a cutpoint between the modes of that distribution. It also appears that people do discretize cleanly, rather than create a probabilistic threshold; DE was uniformly lower than IE, indicating that, even when subjects did not completely recover the underlying distribution, they were attempting to split the feature cleanly. The next chapter looks at the question of the form of the cutpoint more closely.

Chapter 4

Further Analysis: Modeling

As explained in Section 2.2, there are two possibilities for the type of cutpoints people might use: the cutpoints may be either probabilistic or exact. Section 1.2 noted that a great deal of research shows human cognition to be probabilistic, which leads to a strong a priori intuition that discretization will be probabilistic also. However, research in categorical perception (e.g. Harnad, 1987) has shown that human perception divides certain aspects of the environment more sharply than is warranted (e.g., phonemes). This section looks at how the data from the first experiment compared to both a probabilistic and an exact model.

4.1 Statistical Models

The results from Experiment 1 were modeled in order to look more closely at the question of how cleanly subjects discretize. Two possible models were considered: a model which produces an exact discretization, and a model which produces a fuzzy discretization. For the exact model, a cutpoint model was used; for the fuzzy model, a mixture of two Gaussians was used. If the cutpoint model fits the experimental responses more closely, it indicates subjects may be attempting to use an exact discretization, like the discretization used in machine learning. If the mixture model fits the experimental responses more closely, it indicates discretization is similar to other human cognitive processes such as identifying loudnesses (Garner & Hake, 1951). Additionally, the mixture model also indicates subjects are closer to recovering the underlying statistical distribution, and thus there is another level of analysis to carry out: does the model reproduce the underlying distribution, thereby suggesting the subjects are recovering the statistical information in the environment, or is it markedly different?

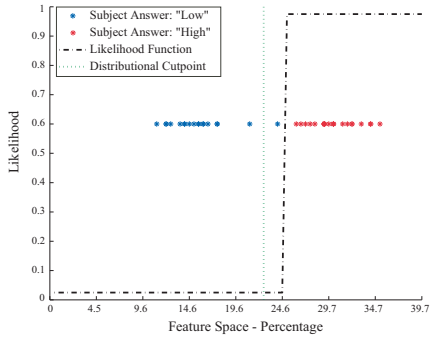

Figure 4.1: Example of cutpoint model with subject's data. Dotted line indicates equal-likelihood point of the original mixture; dashed line indicates likelihood of responding "higher" according to a cutpoint model

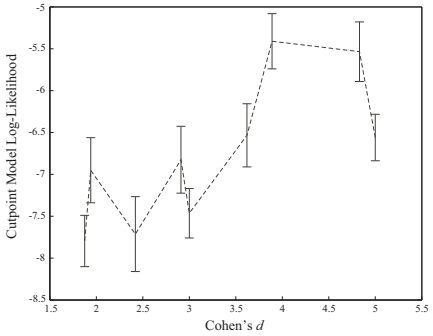

Figure 4.2: Average log-likelihood of cutpoint model as a function of Cohen's d

In order to accurately capture the individual behavior, each block was modeled individually, rather than modeling results averaged over a factor such as subject, Cohen's d, or scedasticity. Experiment 1 yielded a total of 3906 blocks.

4.1.1 Cutpoint Model

A cutpoint model was used to model a clean discretization: any continuous values below a cutpoint C would be labeled as one discretized value (e.g. "narrow" "black"), and any continuous values above the cutpoint would be labeled as the other discretized value. Under the assumption that subjects either respond to a rule or guess randomly, a guessing parameter g was added. Thus, the probability that an item would be put in

the "high" discretization value based on a cutpoint of C, is:

$$p(x \Rightarrow \text{"high"}) = \begin{cases} g & x < C \\ 1 - g & x > C \end{cases}$$

As noted in Chapter 3, subjects' cutpoints did not always match the optimal cutpoint in the underlying data; in order to accurately model the results, therefore, the subject's cutpoint for that block was used. Figure 4.1 shows a single representative block, including the subject's data, the location of the cutpoint at the point of equal likelihoods in the generating distribution, and the cutpoint model that best matches the subject's responses.

In order to calculate the likelihood that the subject's results came from a cutpoint model based at the the given cutpoint, the probabilities of all the items in the block (which are independent) were multiplied together; correctly assigned items had a probability of $p = 1 - g$, and incorrectly assigned items had a probability of $p = g$ (for this model, $g = 0.025$). Thus, the maximum likelihood (that is, if the subject exactly matched the cutpoint model) would be 0.975^n, where n is the number of items in the block. Over all 3906 blocks, 1266 had maximum likelihood; that is, 32.4% of the blocks were discretized completely cleanly. Figure 4.2 shows the average log-likelihood of the cutpoint model, collapsed over Cohen's d, illustrating that, as Cohen's d got higher, the log-likelihood of the cutpoint model tended to go up.

4.1.2 Gaussian Mixture Model

A natural candidate for a "fuzzy" discretizer is a Gaussian mixture model, i.e. a model that assumes the data were generated by a mixture of two Gaussians, and has only to estimate the parameters of the mixture. This model estimates the mixture and determines the Bayes optimal probabilities as to how a particular feature value would be discretized based on the estimated mixture. As the training data were in fact generated from a mixture of two Gaussians , this also serves as "ideal observer," i.e. a model whose assumptions about the environment are correct.

The best-fit mixture model was determined by a brute-force search through a suitably restricted part of the 4-dimensional feature space for two means and two standard

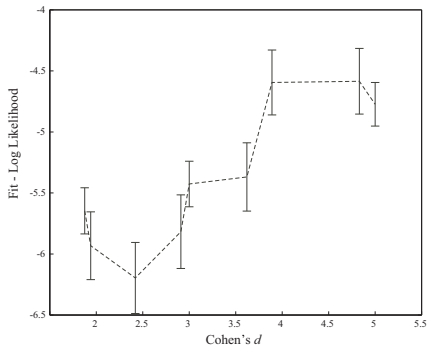

Figure 4.3: Average log-likelihood of the best-fit Gaussian mixture model as a function of Cohen's d

deviations. The two means were fit over separate ranges; each possible mean was restricted to a range from approximately 10% below the associated given mean to approximately 10% above. Because the original homoscedastic standard deviations were different that the original heteroscedastic standard deviations, the range for the two best-fit standard deviations depended on the scedasticity of the original distribution. The heteroscedastic standard deviations ranged separately from just under 1% to just over 16%; the homoscedastic standard deviations ranged separately from just under 1% to approximately 10%. Note that the two estimated standard deviations were not constrained to be either equal or unequal in a particular mixture, meaning that the scedasticity of the estimated mixture did not have to match the scedasticity of given distribution.

To determine likelihood for each combination of means and standard deviations in this range, the search algorithm created two distributions, A and B. The A distribution had a mean closer to 0, while the B distribution had a mean closer to 1. The general likelihood that an item drawn from this summed distribution would be assigned to the "high" discrete value is:

$$p(x \Rightarrow \text{“high”}) = \frac{B}{A+B} \tag{4.1}$$

As in the cutpoint model, a guessing parameter ($g = 0.025$) was added.

$$p = (1-g)\frac{B}{A+B} + \frac{g}{2} \tag{4.2}$$

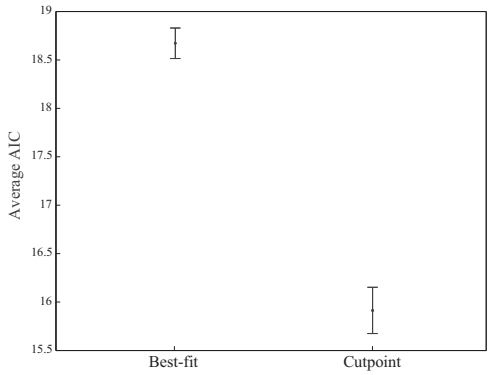

Figure 4.4: Average AIC for the two models over all blocks. Lower AIC indicates a better fit to the data.

Thus, the algorithm calculated the likelihood the subject would respond "high" for each item by taking the the probability at that value, and multiplying the likelihoods (which are independent) to give an overall likelihood for the particular combination of means and standard deviations. The combination of two means and two standard deviations that yielded the highest likelihood for that block were returned. Figure 4.3 shows the average log-likelihood of the best-fit mixture models, collapsed over Cohen's d, illustrating that, as Cohen's d got higher, the log-likelihood of the mixture model tended to go up.

4.1.3 Comparison of Probabilistic Models

The likelihood of the cutpoint model and the likelihood of the best-fit Gaussian mixture model were found for each block. However, because the mixture model has more parameters than the cutpoint model, it is likely to fit the subjects' results more closely. Therefore, rather than comparing the log-likelihoods of the models directly, the two models were compared using the Akaike information criterion (AIC) (Akaike, 1974), as the AIC compensates for different numbers of parameters. The AIC is

$$AIC = -2ln\zeta + 2\rho \tag{4.3}$$

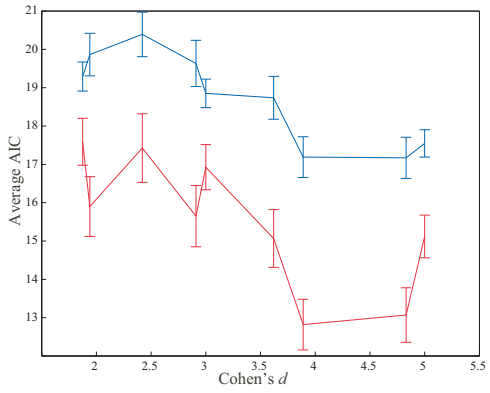

Figure 4.5: Average AIC score as a function of Cohen's d. Lower AIC indicates a better fit to the data. Blue is mixture models, Red is cutpoint models.

where ζ indicates the likelihood of the model and ρ indicates the number of free parameters. When comparing two models, a lower AIC indicates a better-fitting model. The Gaussian mixture model has 4 free parameters (μ_1, μ_2, σ_1, σ_2), while the cutpoint model has none.

Over 3906 total blocks, the cutpoint model was a better fit according to the AIC on 3062 blocks (78.4%). A t-test indicates the average AIC of the two models are different (mean cutpoint model AIC: 15.91; mean mixture model AIC: 18.67; lower AIC indicates better fit. Two-tailed t-test: df: 7810, $p \ll 0.001$) Figure 4.4 shows the average AIC for the two models, collapsed over all blocks. Further analysis comparing the average best-fit mixture AIC over the 844 blocks more closely modeled by a mixture model to the average cutpoint AIC over the 3062 blocks more closely modeled by the cutpoint model shows a larger difference: mean mixture AIC for mixture blocks was 29.64, while mean cutpoint AIC for cutpoint blocks was 10.36 (lower AIC indicates better fit; two-tailed t-test: df: 3904, $p \ll 0.00001$).

The next step in the analysis was to see how Cohen's d affected the models. As demonstrated in Briscoe and Feldman (2006), models' performance may vary depending on the complexity of the training data. Perhaps there is parallel behavior in discretization, such that the Gaussian model is a better representation of human behavior at certain levels of Cohen's d.

However, this is not the case. As can be seen in Fig 4.5, at all levels of Cohen's d, the average AIC for cutpoint models is significantly lower than the average AIC for mixture models. Individual t-tests at each level of Cohen's d confirm this, with p values ranging from 0.01 to below 0.000005. There is also a significant trend down for the AIC for both models (cutpoint model: $F(1, 3904) = 22.3, p \ll 0.001$; Gaussian model: $F(1, 3904) = 30.4, p \ll 0.001$), which reflects the results from the original log-likelihoods indicating that subjects' results are modeled more closely at higher levels of Cohen's d.

Overall, a clean cutpoint distribution modeled the data better than the fuzzy mixture model that more closely resembles the underlying distribution. This suggests that while subjects are influenced by the environmental distribution, they do not recover it ideally: they "hyperdiscretize" the environment by inducing a cut-point more exact than is optimal. This idea will be developed more below. However, there were 844 blocks which, according to AIC were modeled better by a mixture model. The next section takes a closer look at those particular blocks.

Gaussian Blocks

Out of 3906 blocks, 844 were modeled better by a mixture model than by a cutpoint model. Closer analysis shows that the subjects performed more poorly on these blocks. As would be expected, subjects make significantly more Discretization Errors in blocks best modeled by mixture models; mixture models are expressly constructed to be fuzzier than cutpoint models. However, subjects also made more Ideal Errors in blocks best modeled by the best-fit mixture model (two-tailed t-test, $df = 3904, p \ll 0.0001$; see Fig 4.6), indicating that subjects were better at recovering the underlying parameters when they were able to create an exact cutpoint.

4.1.4 Hyperdiscretization

The data have provided several hints that subjects may draw sharper discretization boundaries than are actually supported by the data, a phenomenon that might be called *hyperdiscretization*. In particular, the idea that subjects hyperdiscretize is strongly suggested by the large proportion of blocks that are modeled better by a clean cutpoint

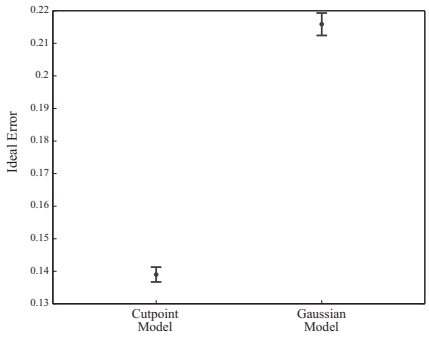

Figure 4.6: Ideal Error by model type

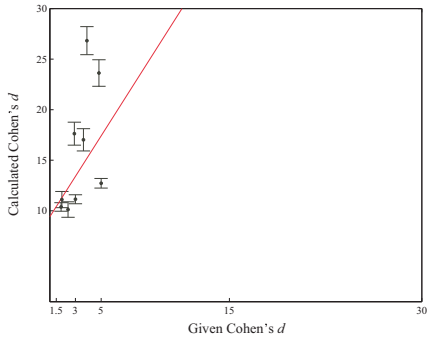

Figure 4.7: Average Cohen's d calculated from the estimated parameters of the best-fit Gaussian mixtures, as a function of the Cohen's d of the underlying distribution. The Red line is the regression line of the data.

model than a more realistic Gaussian mixture model.

Another way of understanding this phenomenon is to return to the Gaussian mixture model, and compare the best-fit models' separations (calculated Cohen's d) with the separation in the source distribution (given Cohen's d). A t-test comparing the two across all 3906 blocks show they are different (two-tailed t-test, $df = 7810, p \ll 0.000001$). Figure 4.7 shows how they compare; as can be seen from the two axes, the Cohen's d values calculated from the best-fit Gaussian models are much higher than the Cohen's d of the source distributions. This indicates the subjects' mental representation of the environmental distribution was substantially "spiker" than it actually was.

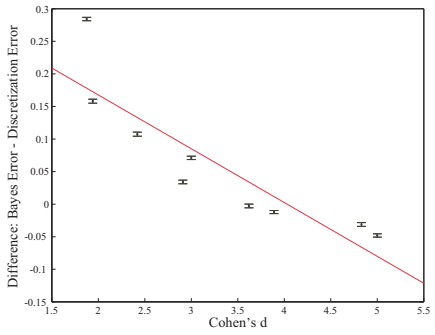

Figure 4.8: Average of Bayes Error - Discretization Error by Cohen's d of the underlying distribution. The Red line is the linear regression that fits the data

Bayes Error

Recall from Section 2.3.1 that Discretization Error indicates how inexact the cutpoint is; thus, a DE lower than would be expected from the generating Gaussian mixture implies hyperdiscretization. The error expected from a Gaussian mixture would be the percentage of the two distributions that fall on the "wrong" side of the cutpoint in the distribution, i.e. the tails. The distributional cutpoint that produces the lowest error rate is the Bayes optimal cutpoint: the likelihood ratio crosspoint. Therefore, this expected error is called *Bayes Error*. Bayes Error is thus the error rate of an ideal observer who knows the parameters of the two underlying Gaussians generating the feature values.

The Difference Score (DS = BE - DE) is thus a measure of the degree of hyperdiscretization. A one-way t-test on this difference score shows that the difference is significantly higher than 0 (one-tailed t-test, $df = 3905, p \ll 0.000001$). That is, subjects' responses have less noise in them than the minimum connected with an ideal observer. This demonstrates subjects are enforcing a cutpoint that is more exact than the ground truth, i.e. that they hyperdiscretize.

Further analysis shows the Difference Score is not constant across Cohen's d. Over all 3906 blocks of Exp. 1, 2753 of the blocks (70.5%) had a DE lower than the BE. Closer analysis shows that there is a significant downward trend in the DS, indicating

DE gets closer to BE as Cohen's d goes up ($F(1, 3904) = 6875, p \ll 0.0001$; see Fig. 4.8). However, only at the highest levels of given Cohen's d does the average difference drop significantly below 0.

4.2 Summary of Modeling

Comparing exact and fuzzy models of the individual blocks from Experiment 1 indicated that subjects attempt to produce an exact cutpoint between discrete feature values, unlike what might be expected considering other cognitive behaviors. The cutpoint model outperforms the mixture model in 78% of the blocks, and the error rates for blocks best modeled by mixture models are significantly higher than those modeled by cutpoint models. Additionally, the Cohen's d calculated from the best-fit mixture models is markedly higher than the given Cohen's d. Finally, comparison of the subjects' Discretization Error to the error inherent to the underlying distribution (Bayes Error) shows that at low and moderate levels of Cohen's d, subjects hyperdiscretize.

Having demonstrated that humans discretize cleanly, the next question of interest is: what else might affect discretization? In particular: will the various values of a second feature influence the discretization of a first? And if so, how?

Chapter 5

Experiment 2 - Interaction of Two Features

The experiments so far have considered discretization when one feature's values vary. With multiple features, several new questions arise, concerning how the features relate. One natural question is whether modal separation along one feature can influence discretization along another. As explained in Section 1.1.1, machine learning theorists have noted that discretizing one feature without respect for another may miss important information in the data (Bay, 2001). However, category research has indicated that human subjects find it much easier to work with one feature. For example, an active research thread explores why subjects prefer to use one feature to create categories instead of a cluster of features that produce family resemblance categories (e.g. Ahn & Medin, 1992; Diaz & Ross, 2006), and other research has demonstrated that subjects are more accurate in learning a boundary between categories when that boundary line is perpendicular to one feature (e.g. Alfonso-Reese, Ashby, & Brainard, 2002). Therefore, the second experiment tested whether the discretization of one feature would be affected by the distribution of a second.

5.1 General Design - Experiment 2

As in Experiment 1, Experiment 2 tested how subjects discretized single features with two modes. Recall that there were three parameters of interest in the first experiment: location in feature space; separation; and number of items.

Because this experiment is primarily interested in the effect of one feature on another, the number of locations was reduced to two, one on either side of the midline, to counteract the effect of mean drift.

The separation parameter is a little more complicated, as this experiment involves

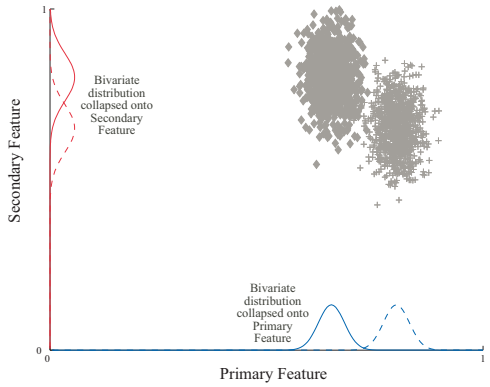

Figure 5.1: Sample bivariate distribution in feature space.

two features; however, in the interest of simplifying comparison with Exp. 1, the underlying distributions for each feature were treated separately. For a given block, when collapsed over the discretized (or *primary*) feature, the stimuli had two modes which had the same distance between means and the same three standard deviations as Exp. 1A; that is, there were three levels of the *primary Cohen's d*. When the stimuli were collapsed over the non-discretized or *secondary* feature, the stimuli again had two modes; the distance between means again matched the distance in Exp. 1A; but there were only two levels of standard deviation. Thus, there were two levels of *secondary Cohen's d*. Figure 5.1 shows a sample set of distributions in feature space.

Finally, the number of items was kept constant at 25 per block.

The three varying factors (separation of the primary feature, separation of the secondary feature, and location in feature space) were completely crossed within subjects, for a total of 24 blocks; a list of the parameters can be found in Appendix A.

5.2 Procedure

The procedure for the second experiment was the same as the first experiment except for two specific details that will be discussed shortly. Each block, the subject studied a group of ellipses on the computer screen, arranged in a grid in random order (Fig 5.2). The subject mentally sorted them into two groups. When the subject felt comfortable about how s/he would assign the ellipses to the two groups, the subject pressed any

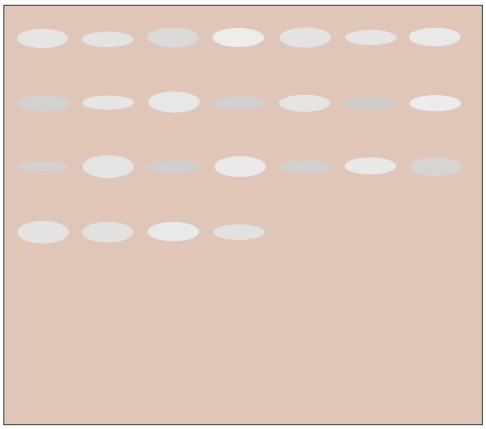

Figure 5.2: Example of "group" screen in Exp. 2

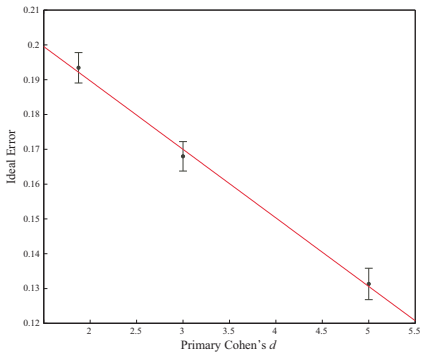

Figure 5.3: Ideal Error as a function of Primary Cohen's d, Exp. 2.

key on the keyboard. Each ellipse was then presented individually in (a new) random order. The individual ellipses were larger than the ellipses on the group screen (where they had been uniformly reduced in size to fit on one screen). The subject was asked to press one key if s/he had decided the ellipse was in one group, and another key if s/he had decided the ellipse was in the other group. After each response, a blank screen appeared and the subject would press the space bar to start the next trial. After the subject classified all the ellipses from that group, a new block with a new set of ellipses drawn from a new mixture would begin. As in the first experiment, after the subject read the instructions but before the start of the experiment, the experimenter reiterated that the subject would be seeing a broad range of ellipses, but that for any given group,

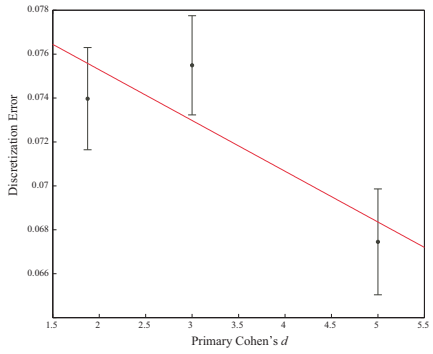

Figure 5.4: Discretization Error as a function of Primary Cohen's d, Exp. 2.

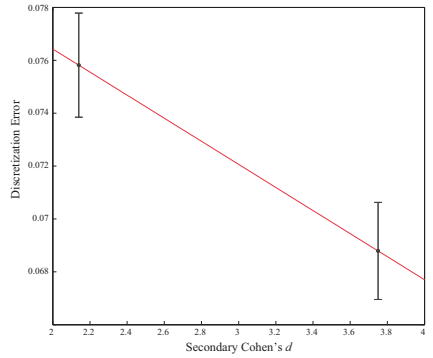

Figure 5.5: DE as a function of Secondary Cohen's d, Exp. 2.

they should decide the sorting based only on the other ellipses in the current group.

As mentioned, there were two main differences between the procedures for Exp. 1 and Exp. 2. The first difference was that, unlike the first experiment, the ellipses in the second experiment varied in both luminance and aspect ratio. The instructions specified which feature the subject would be using to sort the ellipses; this feature was also restated by the experimenter. The second difference between the two experiments is that each subject participated twice. When a subject had finished sorting the blocks on one feature, s/he was given a short break and a new set of instructions; these instructions told the subject s/he would be doing the same task, but s/he was to sort by the other feature. The order in which the subjects sorted by aspect ratio and by luminance was counter-balanced across subjects. Together, the two tasks took approximately an hour.

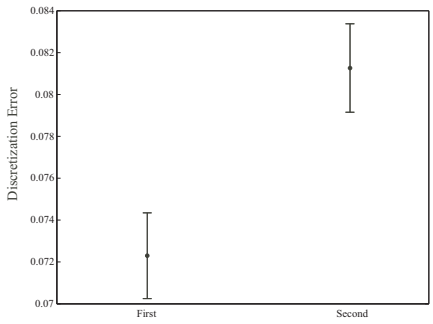

Figure 5.6: DE by task order, Exp. 2.

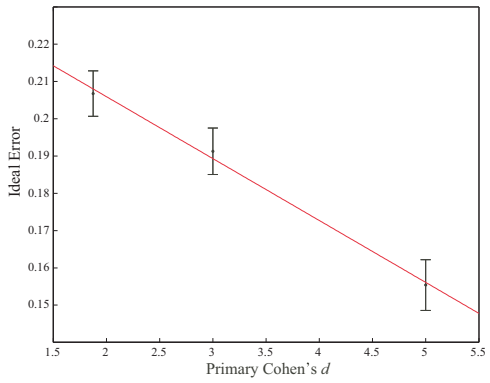

Figure 5.7: IE as a function of Primary Cohen's d – Aspect Ratio blocks only. Exp. 2.

5.3 Subjects

Subjects were 59 members of the Rutgers community who either received class credit or $10 for their participation. One subject was dropped because s/he had not finished the first task by the end of an hour; two subjects were dropped for not doing the tasks correctly, such that half or more blocks in a task were not discretized; four subjects were dropped for having error rates more than three standard deviations from the mean error rates; and one subject was dropped because of a data-recording error: a total of 51 subjects were analyzed.

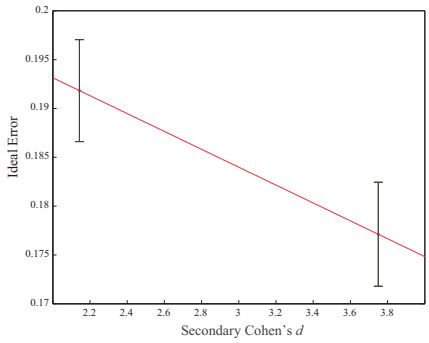

Figure 5.8: IE as a function of Secondary Cohen's d – Aspect Ratio blocks only, Exp. 2.

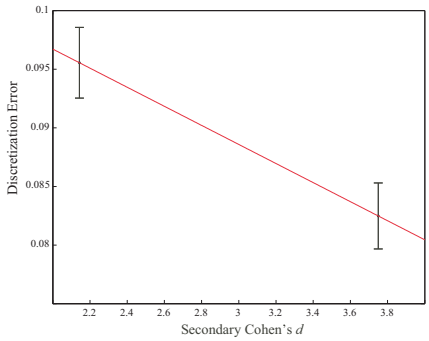

Figure 5.9: DE as a function of Secondary Cohen's d – Aspect Ratio blocks only, Exp. 2.

5.4 Results

Ideal Error and Discretization Error for the primary feature were analyzed to determine how these error rates were affected by both the primary Cohen's d (that is, of the feature the subjects used to sort) and the secondary Cohen's d.

5.4.1 Combined Results

The primary analysis collapses the results over all blocks.

Mean IE was 16.42%, and mean DE was 7.23%. Primary Cohen's d had a significant effect on IE ($F(2, 2446) = 102, p \ll 0.001$; Fig 5.3), and on DE ($F(2, 2446) = 4.9, p < 0.03$; Fig 5.4), such that error rates were lower at higher levels of primary Cohen's d. Cohen's d of the secondary feature did not have a significant effect on IE, but did have

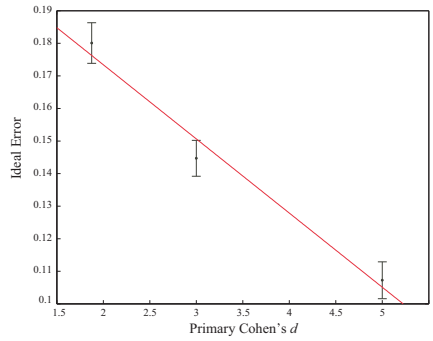

Figure 5.10: IE as a function of Primary Cohen's d – Luminance blocks only. Exp. 2.

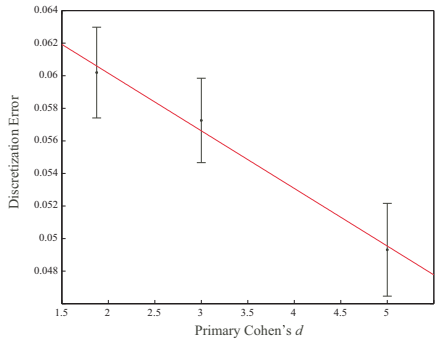

Figure 5.11: DE as a function of Primary Cohen's d – Luminance blocks only, Exp. 2.

a significant effect on DE ($F(2, 2446) = 6.81, p < 0.01$; Fig 5.5), such that DE was lower at the higher level of secondary Cohen's d. There was no significant interaction between the Cohen's d of the two features. Order of task did not have an effect on IE, but did have a significant effect on DE, such that in the second task, discretizations were generally fuzzier ($F(1, 2446) = 7.86, p < 0.01$; Fig 5.6).

Closer analysis showed that there was a significant difference between the error rates for the two features. Mean IE is 14.4% when luminance is the primary feature and 18.45% when aspect ratio is primary (t-test, two-tailed: df: 2446, $p \ll 0.001$); mean DE is 5.56% when luminance is primary, and 8.9% when aspect ratio is primary (t-test, two-tailed: df: 2446, $p \ll 0.001$). Because of this difference, the results were separated by which feature was primary and re-analyzed to determine if averaging was

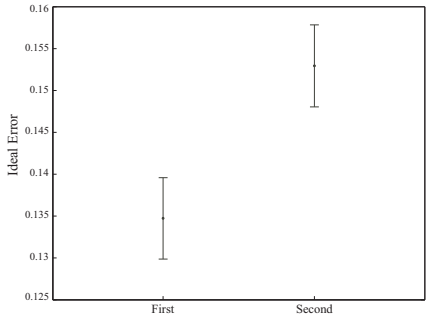

Figure 5.12: IE as a function of whether sorting by luminance was the first or second task – Luminance blocks only. Exp. 2.

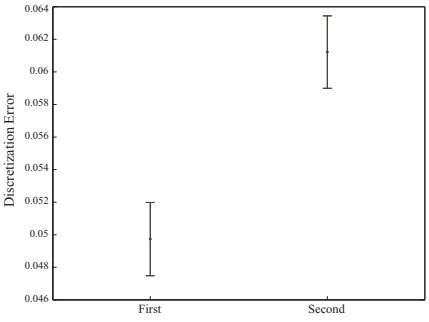

Figure 5.13: DE as a function of whether sorting by luminance was the first or second task – Luminance blocks only, Exp. 2.

smoothing out results in the data.

5.4.2 Aspect Ratio Results

Overall mean IE and DE are reported above. The primary Cohen's d had a significant effect on IE such that error rates were lower at higher levels of primary Cohen's d ($F(2, 1222) = 33.99, p \ll 0.001$; see Fig 5.7), but did not have a significant effect on DE. The secondary Cohen's d had a significant effect on IE ($F(2, 1222) = 3.89, p < 0.05$; see Fig 5.8) and on DE ($F(2, 1222) = 10.03, p < 0.005$; see Fig 5.9), such that error rates were lower at the higher level of secondary Cohen's d. The order of tasks did not have a significant effect on the error rates for Aspect Ratio.

5.4.3 Luminance Results

Overall mean IE and DE are reported above. The primary Cohen's d had a significant effect on IE ($F(2, 1222) = 76.89, p \ll 0.001$; Fig 5.10) and on DE ($F(2, 1222) = 6.74, p < 0.05$; see Fig 5.11), such that error rates were lower at higher levels of primary Cohen's d. Secondary Cohen's d did not have a significant effect on either IE or DE. Sorting order had an effect on IE ($F(1, 1222) = 6.96, p < 0.01$; see Fig 5.12) and DE ($F(1, 1222) = 13.18, p < 0.001$; see Fig 5.13), such that when sorting by luminance was the second task, subjects were less accurate according to both measures.

5.5 Discussion, Experiment 2

The second experiment attempted to determine if other aspects of the environment would affect discretization of a feature. More specifically, it tested whether subjects' ability to discretize one feature would be affected by the distribution of a second. Overall, the distribution of the secondary feature neither improved nor hampered subjects' ability to recover the underlying distribution of the primary feature, but it did affect how cleanly the subjects divided the items. However, this was primarily due to results from aspect ratio as the primary feature; when luminance was the primary feature, the distribution of the aspect ratio feature values did not have a significant effect on subjects' ability to discretize, possibly because the luminance error rate was already comparatively low.

Bay (2001) has argued that a key issue in discretization for knowledge discovery in machine learning is assuring that the discretization does not cover up any interesting patterns in the data. As it is discussed here, the principle behind human discretization is similar: the discretization should capture the useful information in the environment. Some of this useful information will be in the interdependencies of various features. As described in Section 1.1.3, categorization research has shown people can notice and use co-occurances of features in categorization; this experiment examined if people can use co-occurances of feature values to improve discretization. The results, although not conclusive, suggest people may be able to notice categorical distinctions among sources

in the world when they are meaningful. However, further work is needed to determine if the difference between the aspect ratio results and the luminance results is an artifact, or if the effect of the secondary feature does depend on the relative salience of the features.

Chapter 6

Discussion

The original questions of this thesis concerned human discretization. Do humans use information in the environment to discretize? Is their discretization rational or ideal, or not? How cleanly do they discretize?

6.1 Environmental Influences

One of the two general questions this thesis attempted to answer was whether people, when asked to discretize a continuous feature, will use environmental information to determine a discretization that preserves the maximum amount of information.

Both experiments showed that subjects are influenced by various aspects of the environment, as demonstrated particularly by the repeated negative correlation between Cohen's d and error rates. Experiment 1 indicated that subjects will use the distributional information inherent in a collection of feature values to determine a cutpoint. If there is clearly more than one mode in the distribution of feature values, subjects will put cutpoints between the modes. However, if the distribution is comparatively unimodal, subjects' discretizations will be neither as clear nor as accurate relative to the underlying sources of the mixture distribution.

Experiment 1 also suggested that subjects are influenced by the overall range of possible feature values; specifically, when placing cutpoints, subjects showed a preference for the mean of the feature space rather than the ends of the feature space. Indeed, when shown collections of items selected far from the overall mean of the feature space, subjects moved their cutpoints further from the environmental cutpoint than when the items were closer to the overall mean of the feature space. This mean drift indicates that humans remain aware of possible feature values, even if those values are not of

interest at the moment.

Experiment 2 replicated the initial results from Experiment 1, additionally showing that subjects will use the distributional information about a feature to place cutpoints even when other features are also varying. Exp. 2 also provided possible initial evidence that when discretizing one feature, subjects may use the other features of the objects. Subjects' discretizations of aspect ratio were cleaner when the luminance distribution had a clearer separation between modes than when the luminance distribution was more unimodal. Although this result was not replicated when luminance was the primary feature, the possibility of the effect being universal is not ruled out. Luminance error rates were lower than aspect ratio error rates to start; thus, there may be a floor effect preventing the interaction of the secondary feature. Additionally, anecdotal comments from subjects indicated luminance was more salient than aspect ratio; perhaps the effect depends on the relative saliences of the two features.

6.2 Clean Discretizations

The other question this thesis attempted to answer was what kind of cutpoints do subjects use when discretizing: are they clean cutpoints that always assign a particular continuous value to one discrete value, or are they fuzzy cutpoints, such that continuous values are assigned to discrete values at some probability?

Detailed analyses of Experiment 1 indicate that subjects prefer a clean cutpoint; even in cases when they cannot cleanly discretize the space, they hyperdiscretize, such that continuous values near the cutpoint are assigned to the associated discrete value with higher probability than occurs in the underlying distribution. Nearly 80% of the individual blocks from Experiment 1 were modeled more closely by a cutpoint model than by Gaussian mixture model, with more than 30% of the blocks showing perfect discretization. The average Cohen's d derived from best-fit Gaussian mixture models was uniformly higher than the Cohen's d of the underlying distribution. Finally, Discretization Error rates for blocks with low and moderate levels of Cohen's d were significantly lower than the Bayes Error rate calculated from the underlying distribution, indicating

subjects' discretizations were cleaner than the underlying distributions.

6.3 Further Work

Several questions remain open. Experiment 2 presented only minimal evidence that other features can influence discretization; a more complex task, with a larger range of separations and more complex correlations between features will help tease out how one feature influences another. Further research might look at how the complexity of the task and the salience of the secondary feature interacts.

Another open question concerns the number of discrete values. This work focused on discretization into two; how does the number of "bins" in the discretization interact with the separation? Alternatively, when subjects are presented with a group of items and told to sort however they like, will the number of bins they create match the number of modes in the generating distribution? Finally, will the number of modes influence the mean drift? Perhaps subjects showed a preference for the mean of the overall feature space in these experiments because they were dividing the space in two.

Yet a third avenue of inquiry, inspired by the results from Experiment 1 concerning the number of items, involves memory and discretization. A specific task might be to determine if subjects' discretizations are more exact if they are allowed to see all the items during the entire sort.

Chapter 7

Conclusion

There are many continuous features in the natural world that humans must find some way to use. Discretization is one way to reduce the information load without reducing the utility. This thesis has taken the first steps toward learning about human discretization. There are many questions that can be asked about human discretization; this thesis focused on whether humans discretize cleanly, as opposed to using a probabilistic boundary, and on whether humans are influenced by the frequency data available in the environment. Overall, the answer to both questions is "yes." All the experiments, taken together, indicate that humans do use environmental data. As demonstrated by the mean-drift of the cutpoints, and by the influence of the second feature on the first, even information that is not important to the task can have an influence. Additionally, the differences between the Discretization Error and the Bayes Error, combined with the Cohen's d calculated from best-fit models and other results from statistical modeling, indicate that humans prefer a sharp division between discretized feature values.

Taken together, this research opens the door to understanding a fundamental and understudied aspect of symbolic cognition: the process by which discrete symbolic variables are created, and the way they map on to the environment.

Appendix A

Tables

Table A.1: Experiment 1A, Mean Pairs (Homoscedastic).

μ_1	σ_1	μ_2	σ_2	Cohen's d
0.15	0.03	0.3	0.03	5
0.15	0.05	0.3	0.05	3
0.15	0.08	0.3	0.08	1.875
0.2	0.03	0.35	0.03	5
0.2	0.05	0.35	0.05	3
0.2	0.08	0.35	0.08	1.875
0.3	0.03	0.45	0.03	5
0.3	0.05	0.45	0.05	3
0.3	0.08	0.45	0.08	1.875
0.55	0.03	0.7	0.03	5
0.55	0.05	0.7	0.05	3
0.55	0.08	0.7	0.08	1.875
0.65	0.03	0.8	0.03	5
0.65	0.05	0.8	0.05	3
0.65	0.08	0.8	0.08	1.875
0.7	0.03	0.85	0.03	5
0.7	0.05	0.85	0.05	3
0.7	0.08	0.85	0.08	1.875

Table A.2: Experiment 1B, Mean/Standard Deviation Pairs (Heteroscedastic).

μ_1	σ_1	μ_2	σ_2	Cohen's d
0.25	0.04	0.45	0.11	2.42
0.25	0.04	0.45	0.14	1.94
0.25	0.04	0.55	0.11	3.62
0.25	0.04	0.55	0.14	2.91
0.25	0.04	0.65	0.11	4.83
0.25	0.04	0.65	0.14	3.89
0.4	0.11	0.8	0.04	4.83
0.4	0.14	0.8	0.04	3.89
0.5	0.11	0.8	0.04	3.62
0.5	0.14	0.8	0.04	2.91
0.6	0.11	0.8	0.04	2.42
0.6	0.14	0.8	0.04	1.94

Table A.3: Experiment 2, Mean/Standard Deviation Pairs (Two Features).

Primary Feature				Secondary Feature			
μ_1	σ_1	μ_2	σ_2	μ_1	σ_1	μ_2	σ_2
0.2	0.03	0.35	0.03	0.2	0.04	0.35	0.04
0.2	0.03	0.35	0.03	0.2	0.07	0.35	0.07
0.2	0.05	0.35	0.05	0.2	0.04	0.35	0.04
0.2	0.05	0.35	0.05	0.2	0.07	0.35	0.07
0.2	0.08	0.35	0.08	0.2	0.04	0.35	0.04
0.2	0.08	0.35	0.08	0.2	0.07	0.35	0.07
0.2	0.03	0.35	0.03	0.65	0.04	0.8	0.04
0.2	0.03	0.35	0.03	0.65	0.07	0.8	0.07
0.2	0.05	0.35	0.05	0.65	0.04	0.8	0.04
0.2	0.05	0.35	0.05	0.65	0.07	0.8	0.07
0.2	0.08	0.35	0.08	0.65	0.04	0.8	0.04
0.2	0.08	0.35	0.08	0.65	0.07	0.8	0.07
0.65	0.03	0.8	0.03	0.2	0.04	0.35	0.04
0.65	0.03	0.8	0.03	0.2	0.07	0.35	0.07
0.65	0.05	0.8	0.05	0.2	0.04	0.35	0.04
0.65	0.05	0.8	0.05	0.2	0.07	0.35	0.07
0.65	0.08	0.8	0.08	0.2	0.04	0.35	0.04
0.65	0.08	0.8	0.08	0.2	0.07	0.35	0.07
0.65	0.03	0.8	0.03	0.65	0.04	0.8	0.04
0.65	0.03	0.8	0.03	0.65	0.07	0.8	0.07
0.65	0.05	0.8	0.05	0.65	0.04	0.8	0.04
0.65	0.05	0.8	0.05	0.65	0.07	0.8	0.07
0.65	0.08	0.8	0.08	0.65	0.04	0.8	0.04
0.65	0.08	0.8	0.08	0.65	0.07	0.8	0.07

References

Ahn, W.-K., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science*, *16*, 81-121.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716723.

Alfonso-Reese, L. A., Ashby, F. G., & Brainard, D. H. (2002). What makes a categorization task difficult? *Perception & Psychophysics*, *64*, 570-583.

Alvarez, G., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Science*, *106*, 7345-7350.

Anderson, J. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409-429.

Bay, S. D. (2001). Multivariate discretization of continuous variables for set mining. *Knowledge and Information Systems*, *3*, 491-512.

Berger, C., & Hatwell, Y. (1996). Developmental trends in haptic and visual free classifications: Influence of stimulus structure and exploration on decisional processes. *Journal of Experimental Child Psychology*, *63*, 447-465.

Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 458-475.

Briscoe, E., & Feldman, J. (2006). Conceptual complexity and the bias-variance tradeoff. In *Proceedings of the conference of the cognitive science society* (pp. 1038–1043).

Clapper, J., & Bower, G. (1994). Category invention in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 443-460.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (second edition).*

Lawrence Erlbaum.

Cordes, S., Gelman, R., Gallistel, C., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*, 698-707.

Crawford, L., Huttenlocher, J., & Hedges, L. (2006). Within-category feature correlations and bayesian adjustment strategies. *Psychonomic Bulletin & Review*, *13*, 245-250.

Diaz, M., & Ross, B. (2006). Sorting out categories: Incremental learning of category structure. *Psychonomic Bulletin & Review*, *13*, 251-256.

Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. In *International conference on machine learning* (p. 194-202).

Fayyad, U. M., & Irani, K. B. (1993). Multi-interval discretization of continuous-valued attributes for classification learning. In *Proceedings of the 13th international conference on artificial intelligence* (pp. 1022–1027). Choambery, France: Morgan Kauffman.

Feldman, J. (2009). *Symbolic representation of probabilistic worlds.* (Manuscript, Rutgers University)

Feldman, J., & Richards, W. A. (1998). Mapping the mental space of rectangles. *Perception*, *27*, 1191–1202.

Field, D. J. (1989). What the statistics of natural images tell us about visual coding. In B. E. Rogowitz (Ed.), *Society of photo-optical instrumentation engineers (spie) conference series* (Vol. 1077, p. 269-276).

Foard, C., & Kemler Nelson, D. (1984). Holistic and analytic modes of processing: The multiple determinants of perceptual analysis. *Journal of Experimental Psychology: General*, *113*, 94-111.

Friedman, N., & Goldszmidt, M. (1996). Discretizing continuous attributes while learning bayesian networks. In L. Saitta (Ed.), *Proceedings of the thirteenth international conference on machine learning* (p. 157-165). Morgan Kaufman.

Garner, W., & Hake, H. W. (1951). The amount of information in absolute judgments.

Psychological Review, *58*, 446-459.

Grabczewski, K. (2004). Ssv criterion based discretization for naive bayes classifiers. In L. Rutkowski, J. Siekmann, R. Tadeusiewicz, & L. Zadeh (Eds.), *Lecture notes in computer science: Proceedings of the 7th international conference in artificial intelligence and soft computing, june 7-11, 2004* (pp. 574–5796). Springer.

Harnad, S. (1987). Category induction and representation. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (p. 535-565). Cambridge University Press.

Kemler, D., & Smith, L. (1979). Accessing similarity and dimensional relations: effects of integrality and separability on the discovery of complex concepts. *Journal of Experimental Psychology: General*, *108*, 133-150.

Kohavi, R., & Sahami, M. (1996). Error-based and entropy-based discretization of continuous features. In *Proceedings of the second international conference on knowledge discovery and data mining* (p. 114-119).

Kurgan, L., & Cios, K. (2004). Caim discretization algorithm. *IEEE Transactions on Knowledge and Data Engineering*, *16*, 145-153.

Ludl, M., & Widmer, G. (2000). Relative unsupervised discretization for association rule mining. In D. Zighed, H. Komorowski, & J. Zytkow (Eds.), *Proceedings of the fourth european conference on principles and practice of knowledge discovery in databases* (p. 148-158). Springer-Verlag.

Malapani, C., & Fairhurst, S. (2002). Scalar timing in animals and humans. *Learning and Motivation*, *33*, 156-176.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: inference and applications to clustering*. New York: Marcel Dekker.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Monti, S., & Cooper, G. (1999). A latent variable model for multivariate discretization.

In *Seventh international workshop on artificial intelligence and statistics* (p. 249-254). Morgan Kaufmann.

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

Richards, W. A., & Bobick, A. (1988). Playing twenty questions with nature. In Z. Pylyshyn (Ed.), *Computational processes in human vision: An interdisciplinary perspective* (pp. 3–26). Norwood, NJ: Ablex Publishing Corporation.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization.* Lawrence Erlbaum.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382-439.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and brain Sciences*, *21*, 1–54.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *3*, 681–696.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*.

Simoncelli, E., & Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193-1216.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*, 214-241.

Smith, J., & Kemler Nelson, D. (1984). Overall similarity in adults classification: The child in all of us. *Journal of Experimental Psychology: General*, *113*, 137-159.

Smith, L. (1989). A model of perceptual classification in children and adults. *Psychological Review*, *96*, 125-144.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327-352.

Wang, K., & Liu, B. (1998). Concurrent discretization of multiple attributes. In H.-Y. Lee & H. Motoda (Eds.), *The pacific rim international conference on artificial intelligence* (p. 250-259). Springer-Verlag.

Ward, T., Foley, C., & Cole, J. (1986). Classifying multidimensional stimuli: Stimulus, task, and observer factors. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 211-225.

Whalen, J., Gallistel, C., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*, 130-137.

Yang, Y., & Webb, G. (2002). Non-disjoint discretization for nave-bayes classifiers. In C. Sammut & A. Hoffman (Eds.), *Proceedings of the nineteenth international conference on machine learning* (p. 666-673). Morgan Kaufmann,.

Vita

Cordelia D. Aitkin

Education

2004-2009 Ph. D. in Psychology
Rutgers, The State University of New Jersey, New Brunswick, NJ

2001-2004 M.S. in Psychology
Rutgers, The State University of New Jersey, New Brunswick, NJ

1989-1993 BA in Mathematics and Theatre
Williams College, Williamstown, MA

Experience

2006-2009 Rutgers University

- Spring 2009, Instructor, Sensation and Perception
- Summer 2007, Instructor, Cognition
- Spring 2007, Instructor, General Psychology
- Summer 2006, Instructor, Cognition

1999-2001 Plainsboro Marketing Group, Data Maintenance Specialist

1998-1999 Accountants on Call, Temporary Bookkeeper

1996-1998 New York Council for the Humanities, Administrative Officer

1995-1996 Second Stage Theatre, Administrative Assistant

1993-1995 Freelance, New York City Theatrical Lighting Technician

Publications

Aitkin, C. D. and Feldman, J. (2006) Subjective complexity of categories defined over three-valued features. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 961-966.