VISUAL CATEGORY LEARNING WITH DIMENSIONALY-SEPARABLE STIMULI: A COMPARISON OF PERFORMANCE BETWEEN PIGEONS AND HUMANS

A thesis submitted in partial fulfilment of the requirements for the Degree

of Doctor of Philosophy in Psychology

in the University of Canterbury

by Mark E. Berg

University of Canterbury

2009

Acknowledgements

This project and many others undertaken throughout the course of my PhD study would not have been possible without the support, thoughtful guidance, and patience of my supervisor Randolph Grace.

Financial support from the Department of Psychology and the University were vital not only to the survival of myself and family but also in the construction and maintenance of the specialized hardware constructed by Glen Lewis.

Thanks to my participants, the faithful H birds and not-as-faithful G birds, may you return to the cathedral ceiling someday to live out the rest of your days with endless grain. The human participants C, J, M, S & T for you willingness to carry on even in the face of extreme boredom.

My family, Sara, Aida, Lucy, Maya & Eddie, I love you.

Author Note

Portions of Chapter 2 have been submitted for publication:

Berg, M. E., & Grace, R. C. Categorization of Multidimensional Stimuli by Pigeons. Submitted to *Journal of the Experimental Analysis of Behavior*.

Some of the results reported in this thesis have been presented at the following conferences:

Berg, M. E., & Grace, R. C. (2007, May). Visual Categorization in Pigeons: Comparing Rule-Based, Information Integration, and Associative Models. International Paper Session, invited presentation at the annual meeting of the Association for Behavior Analysis, San Diego, California

Berg, M. E., & Grace, R. C. (2007, May). Human Performance in Category Learning Tasks. Poster presented at the annual meeting of the Society for the Quantitative Analysis of Behavior, San Diego, California

Table of Contents

List of Tables

Chapter 2, Experiment 1

Chapter 2, Experiment 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter 2, Experiment 1

Chapter 2, Experiment 2

Chapter 3

Chapter 4

Abstract

Understanding how organisms learn perceptual categories on the basis of experience has been an important goal for researchers in a number of subdisciplines of psychology, including behavior analysis, experimental psychology, and comparative cognition. The primary aim of this thesis is to investigate how nonhumans (pigeons) and humans learn to make visual category judgments when stimuli vary quantitatively along two dimensions, particularly when accurate responding requires integration of information from both dimensions. The thesis consists of four chapters and a technical appendix. Chapter 1 is a literature review which provides a broad overview of studies on categorization by nonhumans and humans, as well as specific background for the current research. Chapters 2 and 3 constitute the empirical portion of this thesis. Four experiments are described, using a category task based on the ‘randomization’ procedure developed originally by Ashby and Gott (1988) with human participants and employed in subsequent research by Ashby, Maddox and their colleagues (see Ashby & Maddox, 2005; Maddox & Ashby, 2004, for review). Stimuli were Gabor patches that varied in frequency and orientation. Our primary goals were to determine whether pigeons could respond accurately in an information integration task with dimensionally-separable stimuli, and to compare performances of pigeons and humans.

Chapter 2 reports two experiments with pigeons. Experiment 1 compared performance in two conditions which varied in terms of whether accurate performance required control by both dimensions (“information integration; II) or by a single dimension (“rule based”; RB). Results showed that pigeons learned both category tasks,

with an average percentage of correct responses of 85.5% and 82% in the II and RB conditions, respectively. Although perfect performance was possible, responding for all pigeons fell short of optimality. Model comparison analyses showed that the General Linear Classifier (GLC; Ashby, 1992), which has been proposed to account for category learning in similar tasks with humans, provided a better account of responding in the II conditions, but a unidimensional model that assumed control only by frequency provided a better account of results from the RB condition. Thus results show that pigeons can respond accurately in an information integration task based on dimensionally-separable stimuli. However, analysis of residuals showed that systematic deviations of GLC predictions from the obtained data were present in both II and RB conditions. Specifically, accuracy for one category (A) was an inverted-U shaped function of orientation, whereas accuracy for the other category (B) did not vary systematically with orientation. Results from the RB condition showed evidence of an interaction between frequency and orientation, such that accuracy was higher for orientation values that were relatively low (i.e. close to horizontal) than high (i.e., close to vertical). Experiment 2 compared responding in two RB conditions which differed in terms of whether frequency or orientation was the relevant dimension. Pigeons again responded accurately in the task. Results from the frequency-relevant condition replicated the interaction obtained in Experiment 1, whereas results from the orientation-relevant condition gave no evidence of an interaction.

Chapter 3 reports two experiments which compare performances of pigeons (Experiment 1) and humans (Experiment 2) in category tasks using identical stimuli. In each experiment there were two conditions, both based on the information-integration task in which the range of orientation values was wide or narrow. There were two primary goals. First, we wanted to test whether the inverted-U shaped pattern for Category A accuracy as a function of orientation would be replicated with different pigeons and stimulus values. Second, we wanted to compare responding of pigeons and humans. A secondary aim was to test whether restriction of range would affect control by orientation. Results from the condition with a wide orientation range were similar to those from Chapter 2, and showed that the inverted-U shaped pattern was replicated for both pigeons and humans. When the range of orientation values was narrow, responding for both pigeons and humans was exclusively controlled by orientation. Overall, results for pigeons and humans were similar and suggest that a common process may underlie information-integration category learning in both species.

Chapter 4 provides a summary of the empirical results from Chapters 2 and 3, and shows that the inverted-U shaped pattern of accuracy for Category A as a function of orientation is unanticipated by current models for category learning, such as the GLC, prototype theory, and exemplar theory. A new ‘fuzzy prototype’ model is described which provides a good account of the results and predicts the inverted-U shaped pattern. According to the new model, subjects associate a linear segment in the stimulus space (‘fuzzy prototype’) with one of the category responses. When a stimulus is presented on a trial, subjects are assumed to use an ‘A/Not-A’ decision rule, with the probability of a Category A response determined as a function of the minimum distance of the stimulus from the fuzzy prototype. Possible directions for future research are considered.

The thesis concludes with a technical appendix which describes the experimental chambers, interface hardware, and computer software developed to conduct the research, and a detailed user's manual for the software. The system allows the same control procedure for both human and pigeon experiments, and should be useful for future research on categorization.

INTRODUCTION

An article published in *Science* in 1964 began a tradition of research on concept formation and categorization research with nonhuman animals. In a classic study, Herrnstein and Loveland (1964) trained five pigeons to discriminate between two types of photographic slides projected in a small cutout screen next to the response key in an operant chamber. The images differed in terms of whether or not they contained the image of a person (in various degrees of obscurity). Responses to the key were reinforced with access to grain when the photograph included a person. Performance showed a high degree of discriminative control, with high rates of responding in the presence of the positive, human-present, stimuli and low rates in the presence of human-absent stimuli. The authors concluded that based on the results there was strong “evidence for the existence of the concept of a person”, and that this was supported by evidence that errors increased as the degree of obscurity increased, or when people-associated objects like boats, cars, and homes were shown in the pictures. Significantly, the categorization performance was very good even when the pigeons were shown novel slides that had never been seen before in early training.

The goal of the research described in this thesis is to explore the ability of nonhumans, specifically pigeons, to solve categorization tasks in which the stimuli vary parametrically on one or two dimensions, and to compare their performance with that of humans. The following literature review will cover a wide range of research on categorization by nonhumans, and include descriptions of representative studies. Noticeably various types of stimuli are used in the experiments summarized, each suited to answer a different type of question related to category learning. Overall, with each study reviewed, evidence builds in favor of the case that categorization is a fundamental ability shared by all organisms, and experimental research has shown that nonhumans can perform to a high degree of accuracy in a broad range of tasks and situations. The literature review is organized as follows. First we examine studies which have investigated the ability of nonhuman animals to respond accurately in perceptual, relational, and associative category learning tasks. Next, relevant human category research will be covered, with particular attention to a leading research paradigm – the ‘randomization’ procedure developed by Ashby and Gott (1988) and used extensively in subsequent research by Ashby, Maddox and their colleagues (see Ashby & Maddox, 2005, for review) – which served as the basis for the research in this thesis. In the randomization procedure, stimuli typically vary on two different dimensions, for example the height and width of a rectangle, or the length and orientation of a line segment. Accurate responding can require attention to one or both of the dimensions. Results from this research paradigm have led to the development of theoretical models for human category learning, such as COVIS (COmpetition between Verbal and Implicit Systems; Ashby, Alfonso-Reese, Turken & Waldron, 1998), which assume that humans have at least two different systems for learning visual categories: An implicit or procedural system, which learns through trial-and-error, and an explicit or rule-governed system, which learns by formulating and testing explicit hypotheses that can often be stated in terms of verbal rules.

Specifically, research in this thesis will explore responding of pigeons in the randomization procedure, with particular attention to tasks in which accurate responding requires attention to both stimulus dimensions. This has been termed ‘information integration’ (e.g., Massaro & Friedman, 1990) because presumably subjects must combine the information from both dimensions prior to making a category decision. Thus the major questions we began with were, could pigeons respond accurately in an information integration task using stimuli identical to those used in human research (Gabor patches), and how would their performance compare to that of humans? We were also interested in the question of optimality – that is, the degree to which pigeons and humans would respond at maximal levels of accuracy in these procedures. Here we provide a brief synopsis of the thesis.

Chapter 1 is a literature review that gives an overview of research on categorization by nonhumans. It then describes categorization research with humans which has used the randomization method pioneered by Ashby and Gott (1988), culminating in the multiple systems view (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998). The theoretical background for the randomization method including signal detection theory and general recognition theory (Ashby & Townshend, 1986) is briefly discussed. Finally, research on categorization which has directly compared performance of humans and nonhumans on similar tasks is described.

Chapter 2 presents two experiments that assess pigeons’ ability to learn categories based on two-dimensional Gabor stimuli that had previously only been used with human subjects. Across conditions, the categories differed in terms of either one or two dimensions. The data are then compared to results with humans in a similar research paradigm, and the current leading model of performance in such tasks, the General Linear Classifier (GLC), is evaluated with respect to the data.

Chapter 3 presents two experiments which compare performance of pigeons and humans on the same task using the same stimuli. Findings from Chapter 1 are examined in more depth, specifically the result that performance was sub-optimal and that there were systematic deviations in the fits of the GLC to the data. In addition, the effects of variation in stimulus range are examined. Chapters 2 and 3 are written as stand-alone manuscripts to be submitted for publication; thus the introductions to these chapters contain some material that overlaps with the literature review.

Chapter 4 explores the implications of the results of Chapters 2 and 3 for theoretical accounts of category learning, and proposes a new ‘fuzzy prototype’ model which performs better than current ones when applied to the data. The question of optimality for both pigeons and humans is also revisited more broadly with respect to previously published studies. Directions for future research are also considered.

The thesis also includes a technical appendix, which describes the specialized hardware and software that were developed to conduct the empirical research reported in this thesis. Special chambers that allow for high resolution touch sensitive display of stimuli are described. Also the software developed specifically to control the customized chambers and display the stimuli is explained. Notably the software allows the same control procedure for both human and pigeon experiments, and should be useful in future comparative research on categorization. A sample program and example version of the software is included in CD form as well as instructions on how to run the example program. Also a software users’ manual is included as a reference.

Chapter 1 – Literature Review

In an introduction to a special issue of the *Journal of the Experimental Analysis of Behavior*, Zentall, Galizio and Critchfield (2002) proposed a classification of the categorization and concept formation literature into three different types of studies. *Perceptual* tasks are those which involve stimuli with small perceptual differences between categories. Often the stimuli may have many shared physical features within each of the categories. For example in the previously-cited study by Herrnstein and Loveland (1964), pictures were grouped into categories on the basis of presence/absence of a person. Other examples of *perceptual* tasks involve representations of real objects like images of paintings, pictures of objects, or real objects themselves. Also included are representations of moving objects, such as video footage or lines in motion presented on a computer screen. Simple perceptual tasks include categorizing simple shapes that vary in width and height as well. In all of these tasks, the stimulus groupings "from a subject's perspective...bear physical similarity to one another." (Zentall et al., 2002, p. 240), regardless of whether the stimuli are pictures of natural objects or lines and shapes. Examples of these studies are described in detail below.

The other types of studies noted by Zentall et al. are *relational* concept studies which manipulate the specific relations among features, for example same-different discriminations (Cook, Katz, & Cavoto, 1997), and *associative* studies in which categories are defined in terms of groups of stimuli which share common functions (Sidman, 1994). These types of studies will also be described in this review.

Perceptual Categorization

Herrnstein and Loveland's (1964) original results were later replicated by Malott and Siddall (1972), and a significant extension of the research was published by Herrnstein, Loveland and Cable (1976). Their goal was to study the development of 'natural concepts', that is, categories based on natural stimuli. They reported three experiments. Experiment 1 or T contained a library of 1840 stimuli containing pictures with or without trees. Experiment 2 or W had 1760 images with or without water and Experiment 3 or P included images with or without a specific person, who had to be distinguished from other people and objects in the pictures. In all three experiments stimuli sets included images that were easy to discriminate with whole or large parts of a person, tree or water and also more difficult images with only small parts or even similar looking components. For example, one difficult image in the T or tree experiment contained a stalk of celery, or clear plastic banana bags meant to make the W (water) discrimination task more difficult. Overall performance in all of three experiments showed higher response rates in the presence of the target stimuli. For example when the pigeons were shown stimuli from the T experiment their pecking rates increased by 90% from rates measured with stimuli without trees. The W and P experiments both showed average increases in response rate of 79% and were consistent even when the pigeons were shown novel pictures that hadn't been seen before. Herrnstein et al. (1976) argued that because no single feature or set of features could explain the pigeons' performance, the most reasonable conclusion was that they had learned the equivalent of a natural concept.

Herrnstein's pioneering work inspired numerous studies on perceptual categorization and concept formation with nonhuman subjects (see Balsam, 1988; Huber, 2001 for review). For example, several studies by Watanabe and colleagues have been influential. Watanabe, Sakamoto, and Wakita (1995) investigated pigeons' ability to discriminate between impressionist or cubist paintings. In their experiment, pigeons were shown various paintings projected onto a 5 cm by 7 cm screen inside an operant chamber. Stimuli were 10 paintings by Monet and 10 by Picasso. Monet is considered an Impressionist artist and uses a style that includes features like visible brushstrokes, light colors, open composition and an emphasis on light. His paintings are typically of nature or landscape scenes, with or without people in various settings. A representative painting, *Terrace of Saint-Adresse*, is shown below in Figure 1.1.

Figure 1.1 *Terrace of Saint-Adresse* by Monet

By contrast, Picasso's style is characterized as Cubist. Artists working in this style typically analyze natural forms such as people or landscapes and then reduce their forms into geometric parts. Their paintings are often 2 dimensional and lacking in visual perspective, and are also typified by a lack of naturalistic color. An example of a Cubist painting is Picasso's *Girls in Avignon*, shown in Figure 1.2 below.

Figure 1.2. *Girls in Avignon* by Picasso

Ten of the typical paintings from each artist were used as training stimuli and then three new paintings from each artist not contained in the training set were used in a generalization test, after initial training in which work of other Cubist and Impressionist artists were used as stimuli. Results showed that pigeons were able to discriminate not only between the artists' paintings during training, but were able to respond accurately in the generalization test as well. In other words, once pigeons had received the initial training to discriminate between Monet and Picasso, they were able to generalize to different artists who worked within the two styles (specifically, the Impressionists

Cezanne and Renoir, and the Cubists Braque and Matisse). A curious result was that generalization even occurred for images of Picasso's work displayed upside down, but not for upside down paintings by Monet. Watanabe has continued his work along the same lines in a more recent article (Watanabe 2001) with a replication of his 1995 study and also a comparative experiment with human and pigeon subjects, which have yielded similar results.

In addition, Watanabe has investigated pigeons' abilities to distinguish real objects from their photographs (Watanabe 1993, 1997). To conduct these studies, he developed a system for displaying stimuli on a beltway that could be observed behind a viewing window: Either real objects or photographs could thus be presented on the beltway. Using this apparatus, Watanabe has been able to test for discriminative ability and also generalization to novel stimuli, all using actual versus photographed objects and images. Accuracy exceeded a 90% discrimination ratio for conditions that compared both objects and pictures of the same objects, and food and pictures of food, even when novel untrained items and pictures were displayed.

A further demonstration of pigeons' ability to form visual concepts based on complex images is reported in the work of Jitsumori and her colleagues, who have used human faces as stimuli. For example, Jitsumori and Makino (2004) exposed pigeons to video images of human faces, and then tested to determine if they could effectively recognize the images when they were shown from different viewpoints. The authors stated that " One of the most sophisticated invariance operations of the visual system is the recognition of objects from novel views," and this experiment sought to build upon other work with rotated objects by Cerella (1977, 1990), Cook, Katz and Cavoto (1999), Wasserman, Gagliardi, Cook, Kirkpatrick-Steger, Astley & Biederman (1996), and Spetch, Kelly and Reid (1999, 2000). To accomplish this, the authors displayed video footage of human faces rotated around the y-axis. Figure 1.3 shows some examples from the stimuli, reproduced in black-and white.

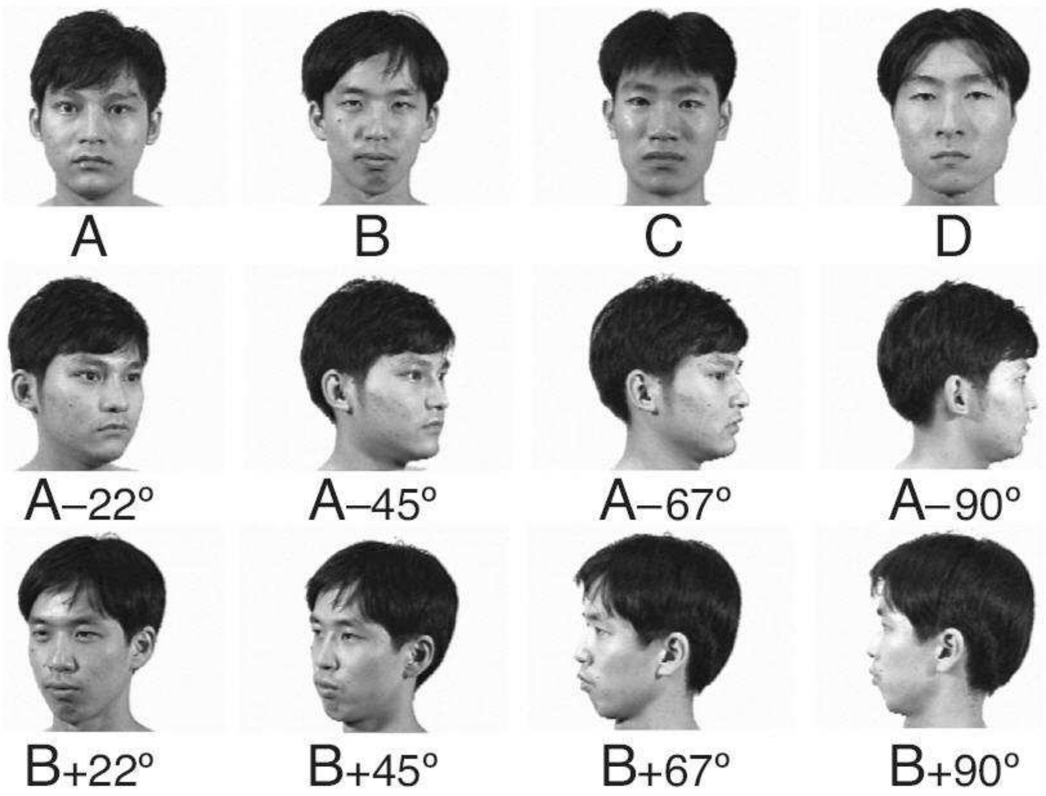


Figure 1. Examples of the black-and-white reproductions of the stimuli. On the top row are the frontal views used for training. The middle and bottom rows are examples of the images from novel viewpoints.

Figure 1.3. Adapted from Jitsumori 2004

In Experiment 1, static views were used in a go/no-go procedure in which pigeons were required to discriminate between positive stimuli like examples C and D and negative stimuli in A and B. After acquiring the discrimination, a generalization test was conducted in which nine different depth orientations were displayed, including the original 0 degree baseline stimuli. Overall, there was a total of 36 different pictures (4 faces x 9 orientations of each face), which were each presented twice per session. Results showed that pigeons' responding generalized from the baseline stimuli across to the novel views of different orientations of the same faces. Jitsumori and Makino suggested that the pigeons were using two possible strategies when responding. The first possibility was that the pigeons based discriminations on the 3-D properties of the photographs and their depth rotation. The other strategy, which the authors concluded was more likely, was that the subjects used multiple 2-D features as discriminative stimuli when generalizing from the training stimuli.

Jitsumori and Makino (2004) reported two additional experiments. In one, they first trained pigeons to discriminate static images of faces. Then they tested the pigeons' ability to generalize from static to dynamically rotating views of the static faces. They tested this by showing rotating video footage of the head of the previously trained image. The results showed that the pigeons failed to generalize from static to dynamic stimuli. In a second experiment the opposite discrimination was trained. First the pigeons were required to discriminate video of dynamically rotating faces, and then were tested for generalization to their static image pairs. Successful generalization between the dynamic and static faces was reported. The authors concluded that "pigeons lump different views together on the basis of similarity judgments, rather than determine whether or not particular views belong to the same 3D object as that previously seen from a different viewpoint."

A recent study by Lazareva, Smirnova and Bagozkaja (2004) compared levels of categorization using photographic stimuli. They conducted daily sessions in which hooded crows were shown various photographs of objects. In half of the sessions (randomly determined), crows were required either to make a basic level categorization by pecking the correct response key (out of 4 keys) depending on whether the photograph showed a car, chair, flower, or a person. Thus there were four categories. In the other half of sessions, the crows were required to make a superordinate level categorization that used only two of the response keys, depending on whether the image was natural or artificial. Examples of each of the four basic level categories are shown in Figure 1.4.

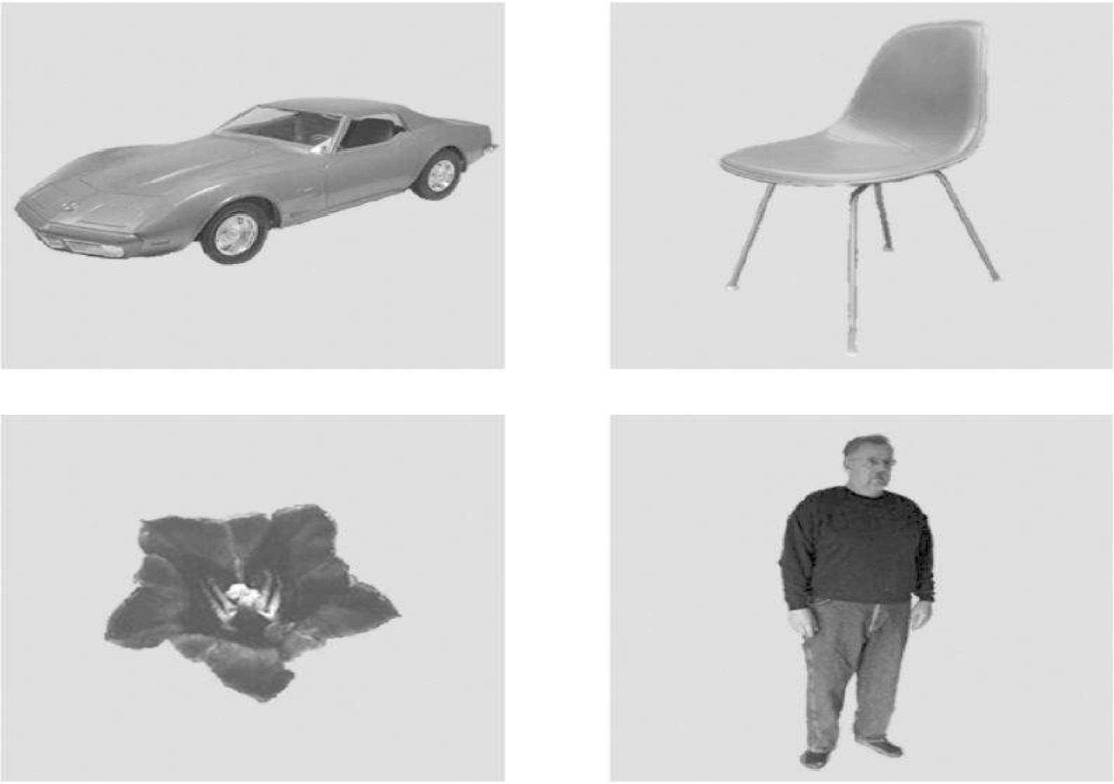


Figure 1. A set of 4 of the 64 stimuli that were chosen from four basic-level categories: cars, chairs, flowers, and people. The complete collection of training and testing stimuli can be seen in color at http://www.psychology.uiowa.edu/Faculty/Wasserman/.

Figure 1.4. Lazareva 2004

The top two images in Figure 1.4 represent artificial or man-made objects in the superordinate category, and the two bottom row pictures represent natural objects. Acquisition of the basic level task was faster than in the superordinate task when viewing the artificial stimuli, but not for the natural stimuli. Results showed that after initial baseline training followed by transfer tests with novel stimuli, crows were able to make both types of category judgments successfully at both the basic and superordinate levels. Based on the reliable transfer of the discrimination, the authors concluded that "such discrimination transfer is the hallmark of conceptual behavior."

Herbranson, Fremouw and Shimp (2002) studied pigeons’ ability to categorize a moving target stimulus. The rationale behind using a moving target was to provide stimuli that represented a more complex and naturalistic situation compared to static images. Pioneering work involving moving targets was conducted by Skinner (1960) with 'Project Pigeon', and subsequent research with dynamic/moving video images has been summarized by Lea and Dittrich (2000). They reviewed various studies which have used video images as stimuli, including discrimination between still and moving images of the same objects (Siegel & Honig 1970), categorization of moving shapes (Emmerton, 1986), and different types of movement and tracking of temporarily invisible objects (Neiworth & Rilling 1987; Neiworth & Wright 1994). The authors conclude that further investigation with dynamic stimuli is important, and that researchers need to avoid trying to address unanswerable questions relating to private events, such as identifying whether or not birds think that video of a conspecific is the same as seeing the subject itself, or whether pigeons perceive color in the same way as humans.

Relational Categorization

A second type of categorization task involves relations between stimuli, such as whether different elements in an array are the *same* or *different*. For example, Wasserman and colleagues have used 16 element visual arrays like the ones shown below in Figure 1.5 (left panel, Wasserman, Hugart & Kirkpatrick-Steger, 1995; right panel, Wasserman, Young & Nolan, 2000).

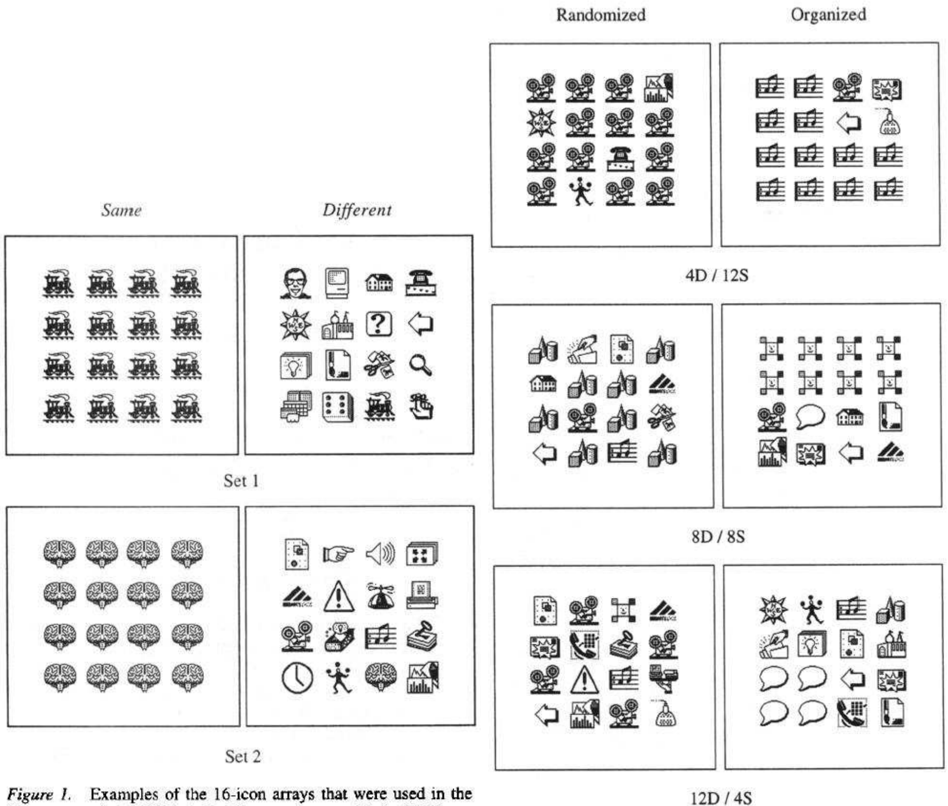


Figure 1. Examples of the 16-icon arrays that were used in the Wasserman et al. (1995) study. These arrays consisted of 16 icons chosen from one of two sets of 16 to create *same* and *different* arrays with these icons randomly located in a 4 × 4 grid.

Figure 2. Examples of the randomized versus organized 4D/12S, 8D/8S, and 12D/4S displays used in Experiments 1 and 2.

Figure 1.5. Wasserman 2000

In the study illustrated in the left panel, Wasserman et al. taught pigeons to discriminate with high accuracy between same and different arrays, as shown in the left and right panels. Pigeons continued to perform at high levels of accuracy when novel same/different displays were presented in a transfer test. Based on the pigeons' ability to discriminate between familiar and novel icon sets, the authors concluded that the pigeons had learned "a general same-different concept" (Wasserman et al., p. 133).

Young and Wasserman (1997) systematically manipulated the amount of variability in stimulus displays comprised of 16 icons, and found that pigeons' ability to discriminate different from same displays increased directly with the entropy of the display, an information-theoretic measure of complexity. They concluded that "the more complex the set, or the more icons contained in the set, the higher the accuracy found in the discrimination results" (Young & Wasserman, 1997). Wasserman, Young and Nolan (2000) conducted a similar study, but systematically manipulated both the variability or number of identical icons contained within the array, and the organization, or location of the similar icons within the array, as shown in the right panel of Figure 6. The authors concluded that both variability (in terms of entropy) and display organization influenced responding: Pigeons were more likely to make a 'different' response when entropy increased, or when spatial organization decreased (i.e., was more random).

Studies by Cook and colleagues have employed a similar methodology of same/different arrays, but have examined the effects of different types of displays. For example, Cook, Katz and Cavoto (1997) used display types that utilized texture, feature, geometric or object stimulus arrays, as shown below in Figure 1.6.

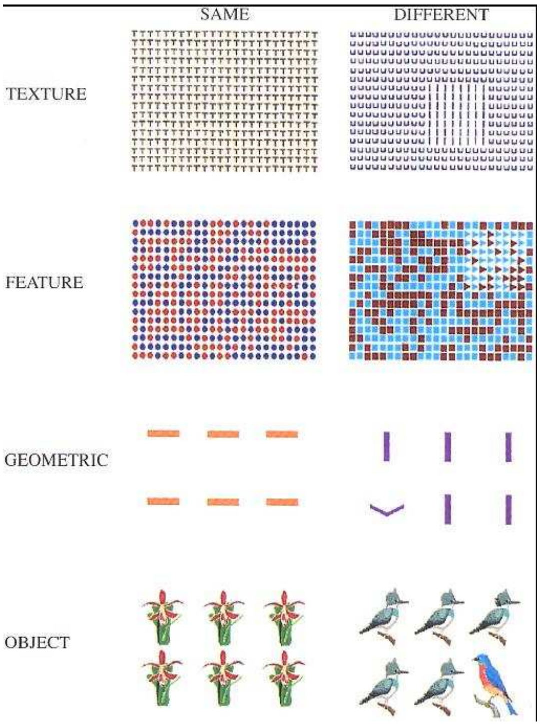


Figure 1.6. Cook 1997, Stimuli examples of each of the four classes with ‘same’ and ‘different’ arrays in the left and right columns.

Figure 1.6 shows examples of stimuli used in Cook et al.’s Experiments 1 and 2. The left column shows ‘same’ examples for each stimulus class, and the right column shows ‘different’ examples. In Experiment 1, pigeons were able to discriminate between same and different stimulus arrays of all classes to a high degree of accuracy. Notably, levels of performance and rate of acquisition were similar for each class. Experiment 2 showed that accurate performance transferred to novel stimuli for each of the four classes. The authors concluded that the results provided strong evidence for the use of a single generalized rule for same/different discrimination utilized by the pigeons.

An alternative approach toward studying relational concept learning was developed by Machado and Cevik (1997), who used a standard three-key operant chamber. They manipulated the relative frequencies of key light flashes during the sample phase of a trial. Then during a subsequent choice phase, the pigeons were reinforced for pecking the key associated with the key light which flashed the least out of the three choices. Overall results indicated that the pigeons were able consistently to perform the task at levels of accuracy significantly above chance as the temporal characteristics of the procedure were varied across conditions. Also of interest was the fact that in one of the experiments the authors tested whether or not the pigeons could generalize the relative frequency of each of the alternatives to ones that were different from those presented in the original sample phase. Results showed that the pigeons were able to generalize accurately to the new and larger overall sample sizes during the choice phase. The authors proposed a model to account for their results, which assumed that as time passed since a stimulus was presented it lost its effectiveness, but that stimuli presented first in the set lost effectiveness more slowly. These assumptions were termed a passive decay and a residual salience process, respectively. Machado and Cevik showed that the model was able to account for both recency and primacy effects that were evident in their data.

Associative Categorization

The third and final type of study we will consider is termed Associative

Categorization. In these tasks, subjects must categorize stimuli in terms of their common consequences or common response requirements. In many cases, the stimuli have no obvious physical similarities or common traits. A popular example of this type of study is stimulus equivalence learning. In equivalence class learning, three criteria must be satisfied for equivalence to be demonstrated (Sidman, 1990): Reflexivity, symmetry and transitivity. For example if A→A and B→B relations are trained, then in test trials emergence of A↔A and B↔B must be shown (reflexivity). In symmetry after A→B, B→C and C→D relations are trained then D→C, C→B and B→A relations must emerge in test trials. Transitivity is demonstrated when after training with A↔B and B↔D relations, A↔D relations emerge in test trials. If the reflexivity, symmetry and transitivity conditions hold, then the stimuli are deemed to form an equivalence class.

Because an ability to learn equivalence classes has been proposed to underlie the development of language in humans (Horne & Lowe, 1996; Sidman, 1992), whether nonhumans are capable of equivalence learning has been an important question. However, research in this area has been controversial: In Sidman and colleagues' original studies and in much subsequent work, results have failed to show that nonhumans have this ability (Sidman & Tailby, 1982; Sidman, Rauzin, Lazar, Cunningham, Tailby & Carrigan, 1982). Sidman (1992) found that only verbally-capable humans showed evidence for transitivity without special training. However, more recently there have been some successful reports of equivalence-like behavior in nonhumans.

Another example of successful equivalence learning with nonhumans was reported by Kastak, Schusterman and Kastak, (2001), who studied California sea lions.

An example of Kastak et al.'s stimuli and procedure are displayed in Figure 1.7.

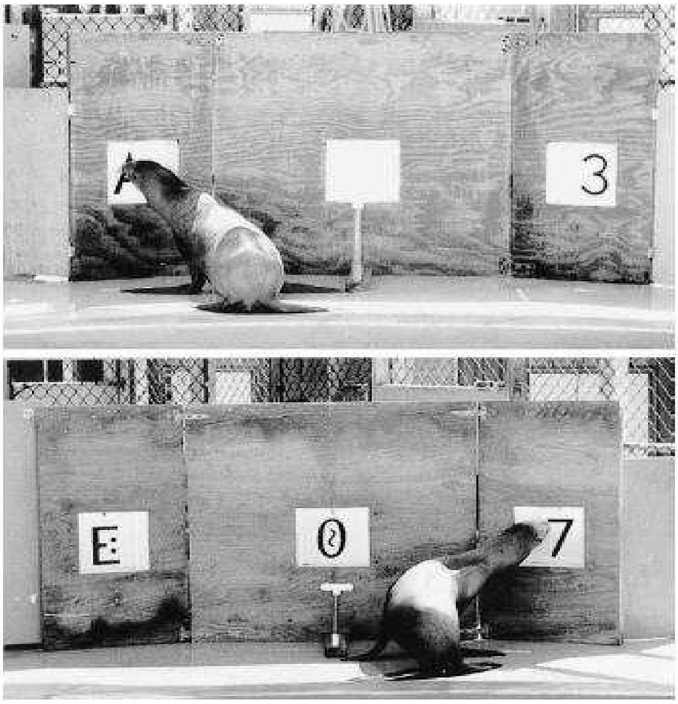


Fig. 1. The top photograph shows sea lion Rio performing a simple discrimination trial. The trial began when the sea lion positioned her head at the stationing bar located in front of the center stimulus box. Following this stationing response, the sliding doors covering the side boxes were opened to reveal Comparison Stimuli A and 3. The sea lion observed the stimuli from her position at the stationing bar until she was signaled by an acoustic cue to make a response. She responded by moving from the stationing bar to touch Stimulus A with her nose. Her correct response was marked by an acoustic tone which signaled that a fish reward would be provided. The bottom photograph shows an example of a conditional discrimination trial. The trial was similar to a simple discrimination, except that following the stationing response and prior to the presentation of the two Comparison Stimuli E and 7, Sample Stimulus 10 was revealed in the center box. Rio's correct selection of Stimulus 7 as the match to the sample was rewarded.

Figure 1.7. Kastak 2001

Experiments were conducted with 2 sea lions named Rocky and Rio. Shown in Figure 1.7 the stimuli were displayed on cards through 3 cut out windows in plywood sheets that stood in front of the subjects. The stimuli were comprised of symbols that had never been trained previously. The sea lions' choices were made by a nose-pointing response, and were observed by a research assistant, who reinforced correct choices with fish. Results showed that the sea lions were able to perform the task successfully, including generalization to the test stimuli and relational transfer between class members. Their results showed that the learned equivalence relationships could be transferred among the three separate experiments that were conducted in the study. Based on the results the authors concluded that if sea lions with no verbal language abilities could learn equivalence relations in agreement with Sidman's (1994) definitions then other non verbal species could possibly learn the same behaviors given the proper training techniques.

Studies with pigeons provide some evidence of equivalence-like performances. For example, Jitsumori, Siemann, Lehr and Delius (2002) reported evidence for transitivity in pigeons using abstract light-emitting diode (LED) patterns as stimuli and a repeated-reversals procedure. The apparatus and stimuli examples are shown below in Figure 1.8.

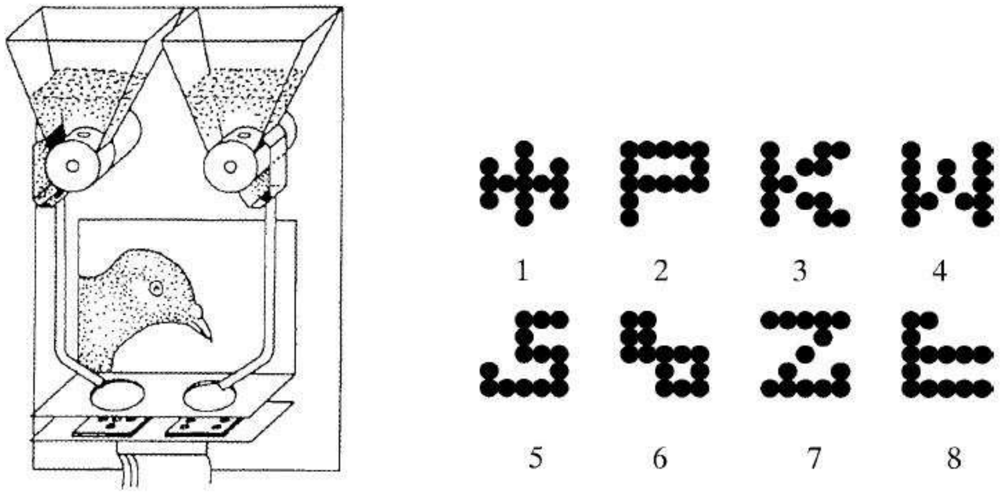


Fig. 1. Left: sketch of the conditioning panels (from Xia, Delius, & Siemann, 1996). During the experimental sessions, they replaced the food troughs of the home cages. Right: the light-emitting diode matrix patterns used as stimuli.

Figure 1.8. Jitsumori et al. 2002

The left panel of Figure 1.8 shows the apparatus used in Jitsumori et al.'s (2002) experiments. During sessions, response keys with LED arrays behind each key were placed on the bottom of the panel where access to food and water was normally provided. Grain dispensers located above the key area were then used for reinforcer delivery. The right panel shows sample stimuli from each of the two sets. The examples in the figure are represented in 4 rows of A, B, C and D stimuli with 2 rows of each of the sets. Jitsumori et al. found that for two of four pigeons, an untrained A↔D relation emerged in transfer tests after A↔B and B↔D relations were trained. These pigeons made an average of 84% correct responses during test (by comparison, the two pigeons that did not successfully demonstrate transitivity made an average of 61% and 64% correct responses).

Frank and Wasserman (2005) showed that pigeons were capable of associative symmetry after having been trained on a successive matching task (i.e., go/no-go discrimination). This contrasted with results of studies using two-alternative symbolic matching tasks in which the comparison stimuli were presented simultaneously (e.g., Lionello-DeNolf & Urcuioli, 2002), which had generally failed to find evidence for symmetry in pigeons (see Lionello-DeNolf, 2009, for review). Recently Urcuioli (2008) has proposed a theory of equivalence class formation that predicts why pigeons show symmetry after successive but not simultaneous matching training. According to this account, pigeons learn the stimulus-temporal compounds that are associated with reinforcement, such that the stimuli together with their temporal locations become members of a class. When different classes share common elements, they tend to merge. Urcuioli (2008) showed that his theory predicted the emergent symmetry in Frank and Wasserman's (2005) procedure. In addition, Urcuioli's theory made a counterintuitive 'anti-symmetry' prediction, in which pigeons would respond more to a non-reinforced symmetric relation if hue-oddity rather than hue-matching had been part of their initial training. This prediction was confirmed (Urcuioli, 2008, Experiment 5). Urcuioli's work is important because it shows how equivalence relations can be the natural result of reinforcement contingencies, identifies conditions under which equivalence relations can be trained in pigeons, and provides a reason for previous failures. However, it should be noted that according to Urcuioli's account, a red keylight that precedes a green keylight is not the same stimulus as one that follows a green keylight; that is, temporal location is intrinsically part of the stimulus (cf. Miller & Barnet, 1993).

An earlier well-known study on stimulus equivalence in pigeons by Vaughan (1988) should also be noted here. Vaughan (1988) used 40 photographic stimuli which were randomly divided into two classes of 20 stimuli each. Pigeons were then trained on a go/no-go discrimination in which stimuli from one class were S+ while those from the other class were S-. After the discrimination was acquired, the stimulus classes were reversed (i.e., S- was now S+ and vice versa). After many reversals, Vaughan showed that on the first session following a reversal, discriminative control for stimuli that had not yet been presented was higher than chance. This suggests that pigeons had learned to treat the different photographs in each class as equivalent. Although Vaughan (1988) claimed that his results showed that pigeons were capable of stimulus equivalence, Hayes (1989) pointed out that because explicit tests for symmetry, transitivity and reflexivity were not carried out, the pigeons' performance could not be considered as demonstrating equivalence. In particular, symmetry and transitivity are emergent relations because they can be obtained without explicit training, whereas one could argue that Vaughan's (1988) results could be traced directly to a history of reinforcement.

Thus, although there have been many failures to show transitivity and symmetry in pigeons, studies have shown that given appropriate training and testing procedures, such as in Frank and Wasserman (2005) and Vaughan (1988), pigeons are capable of performances that resemble equivalence learning.

Summary

Overall, research on categorization with nonhuman animals has shown that a wide variety of species can respond accurately across a range of different tasks, including perceptual, relational, and associative categorization tasks with both naturalistic and artificial stimuli. These results suggest that evolutionary history may have endowed organisms with a common process for category learning. To better understand that process, we next turn to research that has investigated perceptual category learning in which the stimuli and the difference between the categories are defined as simply as possible, in terms of quantitative variation on one or two dimensions. From this perspective, category learning may be viewed as a conditional discrimination, in which the organism learns to make one response in the presence of a stimulus from one category, and another response in the presence of a stimulus from the other category. The classic example of this situation is the signal detection paradigm.

Signal Detection Theory

Signal Detection Theory (SDT) provides a useful starting point for developing a model that describes the process behind category learning. SDT was first developed by researchers studying target detection in radar arrays (Marcum, 1960), and subsequently the concepts were adapted for use in psychophysics research (Swets, 1964). Swets proposed that humans are not just passive information processors but that information is judged actively under uncertain and sometimes difficult conditions. In the standard auditory detection paradigm, either white noise (“noise”) or a tone signal superimposed on white noise (“signal + noise”) is presented to a subject on each trial. The subjects’ task is to identify whether the signal or noise occurred on each trial. According to SDT, the stimuli (noise or signal + noise) are represented with variability in the brain, such that the subjects’ percept can be described in terms of two probability distributions. Sample distributions are shown in Figure 1.9 below. In these examples, the noise is represented by the distribution on the left, and signal + noise by the distribution on the right. The subject is assumed to set a criterion, or decision boundary, such that if the percept sampled on a given trial is to the right of the boundary, the subject responds ‘signal’, and otherwise responds ‘noise’. Specifically, the subject is assumed to compute a likelihood ratio, and respond ‘signal’ if that ratio exceeds a criterion (e.g., 1.0). If the distributions overlap substantially, as in the left pair, accuracy is relatively poor, whereas if the distributions are more separated, as in the right pair, accuracy increases. The standardized distance between the means of the distributions is *d’* or discriminability.

According to SDT, every response has four possible outcomes; hit, miss, false alarm and correct rejection. The subject is assumed to set the criterion so as to maximize the overall accuracy or payoff. Importantly, responding can be biased by differential reinforcement of hits and correct rejections. For example, increasing the payoff for hits and/or decreasing the penalty for false alarms will increase the overall probability of a ‘signal’ response, which corresponds in SDT to shifting the criterion to the left. Conversely, increasing the payoff for correct rejections and/or decreasing the penalty for misses will increase the probability of a ‘noise’ response, which is associated with a shift in the criterion to the right. If the hit rate (i.e., p(‘signal’ | signal) and false alarm rate (i.e., p(‘signal’ | noise) are plotted for a range of criterion values, the Receiver Operating Characteristic (ROC) curve results, which provides a summary of the subjects’ discrimination performance. The lower panel of Figure 9 shows an example of a ROC curve (Heeger, 2003). It is also important to note that SDT is a normative model for detection in that it specifies conditions for optimal responding: the ‘ideal observer’ sets the criterion so as to maximize the overall payoff. In the typical case in which payoffs are equal, this translates into maximizing overall accuracy.

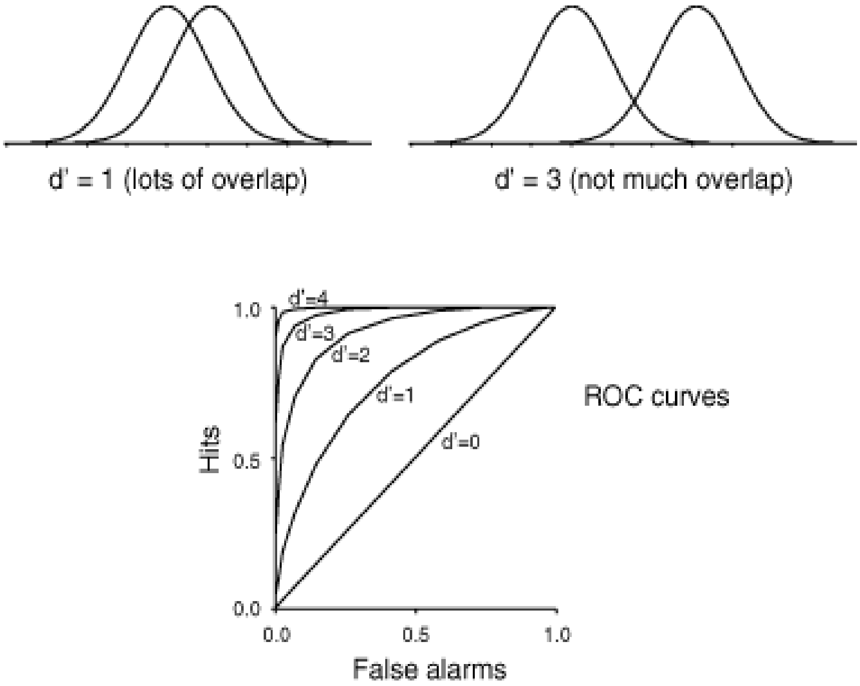


Figure 1.9. Heegar 2003, Signal and ROC curve examples for various d’ values.

SDT has been one of the most influential theoretical frameworks in all of psychology, and has had far-ranging impact in many applied areas (see MacMillan and Creelman, 2005, for review). For example, the area under the ROC curve is commonly used as a measure of discriminative accuracy in applied decision-making contexts such as the prediction of recidivism by violent offenders (e.g., Rice & Harris, 1995). Research has found that responding of nonhumans in detection tasks is similar to that of humans (see Alsop, 1998, for review). Research with nonhumans has emphasized the fundamental reinforcement contingencies present in the detection paradigm. Specifically, the basic detection task can be viewed as two concurrent schedules, each with a different discriminative stimulus: In the presence of ‘signal’, the subject can respond either ‘signal’ or ‘noise’, and similarly in the presence of ‘noise’, can respond ‘signal’ or ‘noise’. This research has led to the development of behavioral models for signal detection (Alsop, 1991; Davison & Tustin, 1978; Davison & Nevin, 1999; White & Wixted, 1999).

General Recognition Theory

An important extension of SDT, with direct relevance for the research reported in this thesis, is the General Recognition Theory (GRT) proposed by Ashby and Townsend (1986). GRT is a generalization of SDT to two or more dimensions. Figure 1.10 below (adapted from Ashby & Gott, 1988) provides a diagram of the two-dimensional case. Stimuli (labeled ‘A’ and ‘B’) are assumed to vary on two dimensions (x and y). On any given trial, the stimulus presented gives rise to a percept which can be represented as a sample from a bivariate normal distribution (upper panel) in a two-dimensional stimulus space. A pair of equal-likelihood contours is shown in the lower panel. If the variances are equal and the covariances are zero (i.e., the dimensions are independent), then the optimal decision bound is the diagonal line shown in the lower panel. That is, the subject can maximize overall accuracy by responding Category A when the percept is above the diagonal line, and Category B when the percept is below the diagonal line. The shapes of the equal-likelihood contours depend on the covariance of the two distributions; as the covariance increases from zero value the shape moves from circular to elliptical. Note that the decision bound is analogous to the criterion in SDT, and represents how an ideal observer should perform in the task.

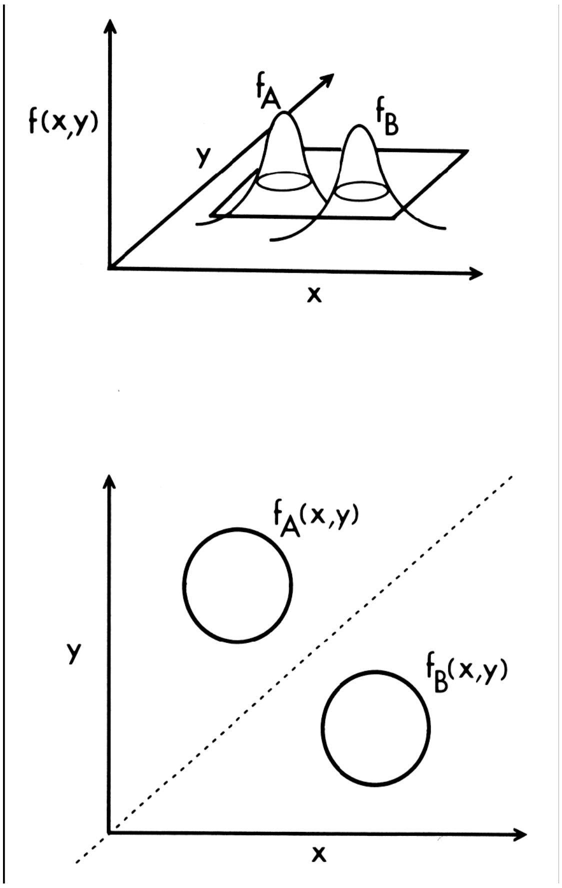


Figure 1.10. Ashby & Gott 1988, Contours of equal-likelihood with decision bounds.

An important aspect of GRT is that the stimulus dimensions are assumed to be separable and independent. In most typical applications of GRT, categories are defined in terms of multivariate normal distributions; Ashby and Townsend (1986) noted that GRT was a prototypical general Gaussian recognition model, and directly related to Thurstone’s (1927) law of categorical judgment.

Ashby and Gott (1988) developed a methodology for empirically investigating situations such as those in Figure 11 known as the Randomization Technique (see also

Ashby & Maddox 1992). The technique is typically utilized in categorization tasks where multiple stimuli values are associated with one of two categories in a ‘many to one’ categorization task. Each stimulus varies on two values, for example the height and width of a rectangle (Ashby & Gott, 1988), the radius and orientation of a circle segment (Ashby & Maddox, 1990), or the frequency and orientation of a Gabor patch (Maddox, Ashby & Bohil, 2003). Categories are represented in terms of bivariate normal distributions. If the variances of the distributions are unequal the ideal decision bound shown by the dotted line in Figure 1.10 would be a quadratic function. For research described in this thesis, the variances for the distributions are equal. When the variances are equal between the categories, a specific model based derived from GRT called the General Linear Classifier (GLC) can be used to examine performance in detail. The essential feature of the GLC is that the decision bound is a linear function in the two-dimensional stimulus space.

GLC Application

The GLC has been used in the majority of categorization work done by Maddox and colleagues (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Maddox & Ashby, 2004; cf. Massaro & Friedman, 1990). In most recent studies, stimuli have been Gabor patches, which are sine-wave gratings modified through a circular Gaussian filter, which vary in terms of frequency and orientation (Yao, Krolak, & Steele, 1995). Usually the stimuli are plotted with Frequency on the x axis and Orientation on the y axis as shown in Figure 1.11. Each stimulus is represented as a point in the two-dimensional space. The different categories are indicated by filled and unfilled symbols.

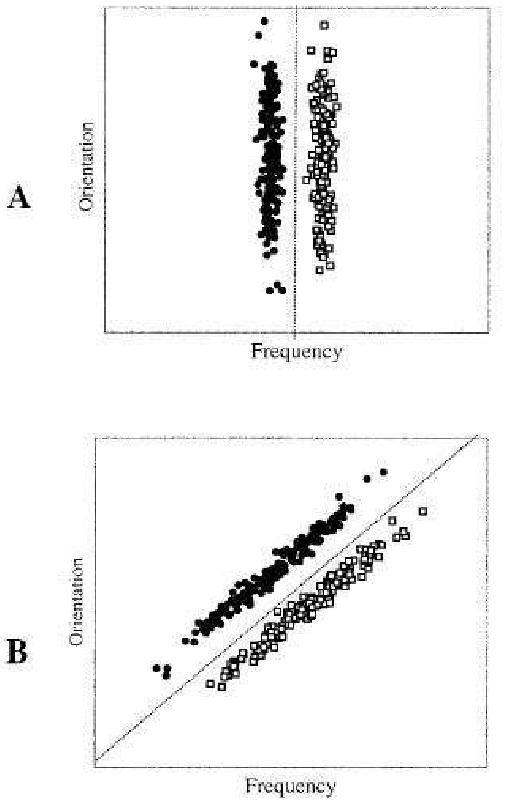
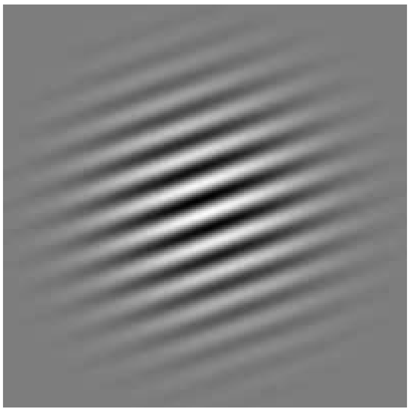
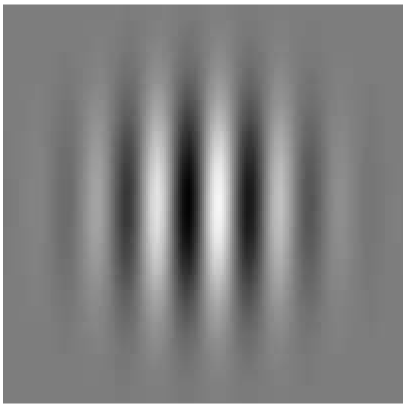


Figure 2. A: Rule-based category structure from Experiment 1. B: Information-integration category structure from Experiment 1. Each circle denotes the spatial frequency and spatial orientation of a Gabor pattern from Category A. Each square denotes the spatial frequency and spatial orientation of a Gabor pattern from Category B. The dotted line in each panel denotes the location of the optimal decision bound.

Figure 1.11. Maddox, Ashby et al. 2003

The top panel represents a situation in which categories differ in terms of frequency (*x*) but not orientation. The vertical line represents a decision bound, such that the subject responds Category A if the frequency is to the left of the line, and Category B if it is to the right. This is a single dimensional or selective attention task, and is also called a Rule Based (RB) task because accurate performance can be described in terms of a simple rule. But if the stimuli and decision bound are rotated by 90 degrees, as in the bottom panel, then accurate performance requires attention to both dimensions. This is a two-dimensional task and is called an Information Integration (II) task. Sample stimuli from the II and RB task are shown in Figure 1.12.





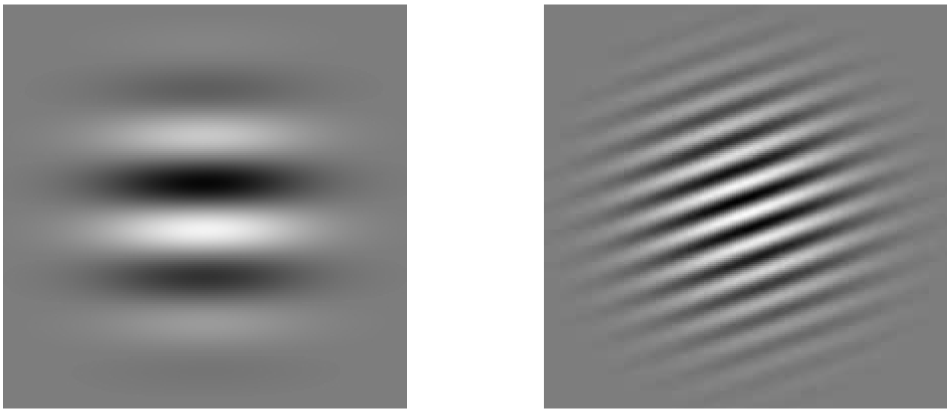


Figure 1.12. Example Gabor images from both the II and RB tasks studied by Maddox, Ashby & Bohil. 2003

The typical trial consists of viewing the stimulus on-screen and then responding with the keyboard to indicate a category A or B response. Subjects generally receive feedback after each trial. It is notable that for both tasks in Figure 11, perfect performance is achievable because the category distributions do not overlap. By contrast, in the original application of the randomization technique shown in Figure 1.11 (Ashby & Gott, 1988), the category distributions overlapped and so 100% accuracy was impossible. In some experiments by Maddox and colleagues, ill-defined overlapping category structures with Gabor stimuli have been used as well. Results of these studies reviewed by Maddox and Ashby (2004) have shown that human subjects are able to perform well in both the Rule Based (RB) and Information Integration (II) tasks.

A particular focus of research using RB and II tasks has been to determine whether performance in these tasks can be dissociated, suggesting that different category learning systems might be responsible for accurate performance. For example, Maddox, Ashby, and Bohil (2003) compared the effects of delayed feedback on performance in RB and II tasks.

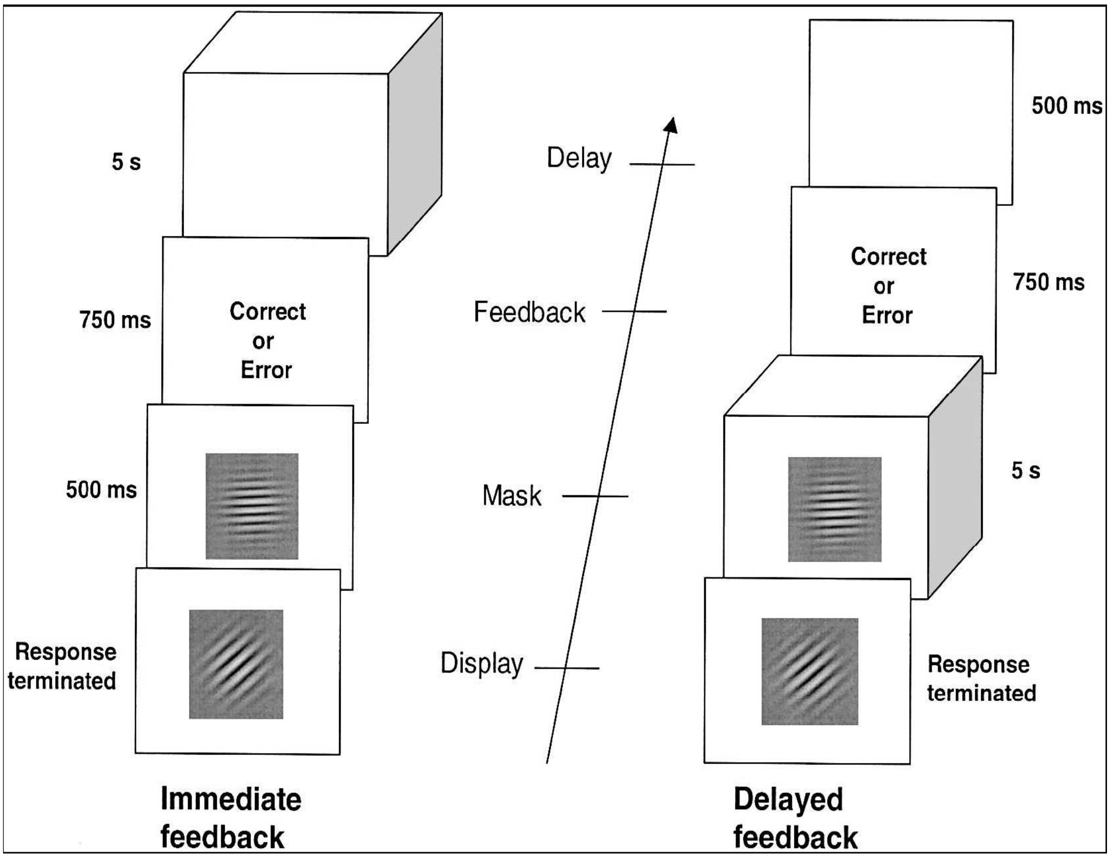


Figure 1.13. Maddox, Ashby et al. 2003, Procedural diagrams from the immediate and delayed feedback conditions.

Figure 1.13 shows the basic procedure that Maddox, Ashby and Bohil (2003) used. In the left of the figure, a trial in the immediate feedback condition is outlined. The stimulus is displayed until the subject responds, and then after a 500 ms masking stimulus, on-screen feedback indicates to the subject a correct or incorrect categorization. The right-hand side of Figure 1.13 illustrates a trial in the delayed feedback condition. The major difference is that the feedback is delayed by 5 s after the masking stimulus. Maddox et al. (2003) found that accuracy was similar for the RB task regardless of whether feedback was immediate or delayed, whereas for the II condition accuracy was reduced in the delayed feedback condition. Also the authors conducted a model comparison analysis to determine the type of strategy used by the subjects. This analysis suggested that for the II task in the delayed feedback condition, subjects were more likely to use a single-dimensional solution (i.e. attend to only one dimension), whereas in the immediate feedback condition they attended to both dimensions.

In other research, Maddox, Ashby and colleagues have shown that various manipulations, such as feedback disruption (Maddox, Ashby, Ing & Pickering, 2004), changes in motor requirements when making responses (Ashby, Ell & Waldron, 2003), and training without feedback (Ashby, Maddox & Bohil, 2002), affected performance on either RB or II tasks, but not both. The procedure utilized in these studies was similar to the one shown in Figure 1.13 (Maddox et al. 2003) with small changes. In the study involving feedback disruption (Maddox et al. 2004), trials either included a long or short feedback processing time allowance. The effects of allowing different amounts of time to process the feedback were different than the result found in the Maddox et al. (2003): Accuracy was reduced in the RB condition when processing time was short, whereas there was no difference in accuracy in the II condition depending on whether processing time was short or long.

Ashby, Ell and Waldron (2003) manipulated the motor requirements during RB and II tasks. In their study, subjects were first given baseline training in order to learn category assignments. Then they were required to either cross their hands over one another or given instructions to ‘mentally’ reverse the key assignments and then press the response keys in order to categorize the stimulus. The result was that accuracy in the II but not RB task decreased in the ‘mentally’ reversed key condition. There was no difference in performance for the II and RB tasks in the hand-crossing condition.

Ashby, Maddox and Bohil (2002) studied the effects of presence or absence of feedback. In the feedback-absent condition, subjects were given a category label and then shown an example from that particular category. The feedback-present condition used an indication of correct or incorrect following the display and subsequent categorization response by the subject. Once again there were differences in the results between either the II or RB tasks. Apart from the overall increase in accuracy in the feedback present condition for the both the II and RB stimuli there was a significant interaction between the two conditions, showing that the presence of feedback had a greater impact on accuracy in the II condition.

Ashby, Maddox and their colleagues have interpreted these findings and others (e.g., Maddox and Filoteo 2001; Maddox, Ashby et al. 2004; Maddox, Filoteo et al. 2004) as evidence that humans have two different systems for category learning as proposed by the COVIS model (COmpetition between Verbal and Implicit Systems; Ashby, Alfonso-Reese, Turken & Waldron, 1998). According to COVIS, one system is verbal, uses explicit reasoning and is under conscious control, and is based neurologically in the cingulated and pre-frontal cortex, whereas the other is a nonverbal, implicit system that depends on reinforcement feedback and is based in the striatum. According to COVIS, the explicit system is used by humans to learn the RB task, whereas the implicit system is used during the II task.

Nonhuman Research with the Randomization Procedure

Research investigating performance of pigeons using the Ashby and Gott (1988) randomization technique has been reported by Herbranson, Fremouw and Shimp (1999). Herbranson et al. showed that pigeons could successfully categorize rectangles of various heights and width dimensions that were displayed on a computer screen. The authors used two overlapping bivariate normal distributions of stimuli with equal zero covariance, similar to Ashby and Gott (1988). The distributions are shown in Figure 1.14 below, which is taken from Herbranson et al. (1999):

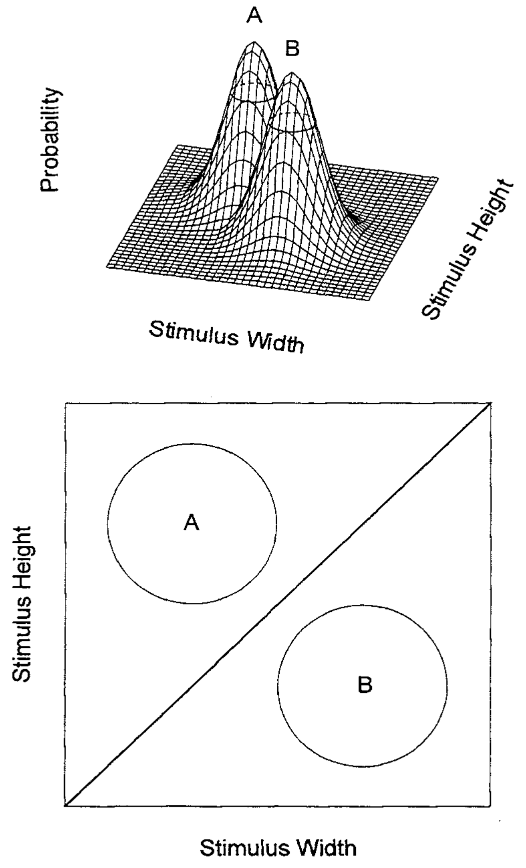


Figure 1.14. Herbranson, Fremouw & Shimp 1999, bivariate normal stimuli distributions with likelihood contours in the top panel and decision line in the bottom.

The two bivariate normal distributions are shown in the upper panel of Figure 1.14, and the equal-likelihood contours are shown in the lower panel. The optimal decision bound is linear, as shown in the lower panel.

Herbranson et al. (1999) studied performance of pigeons in both II and RB tasks, which they described as ‘divided attention’ and ‘selective attention’ respectively. Equal-likelihood contours and sample stimuli from both tasks are shown in Figures 1.15 and 1.16.

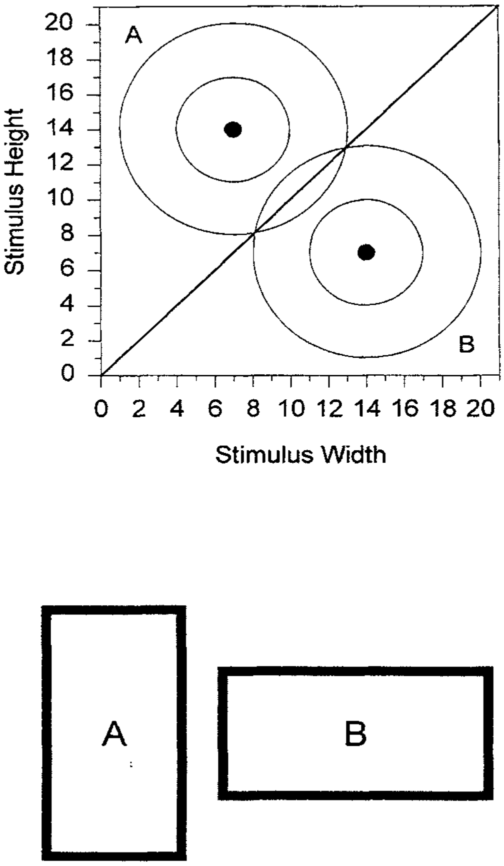


Figure 1.15. Herbranson, Fremouw & Shimp 1999, Stimuli distributions and rectangle examples for the II task.

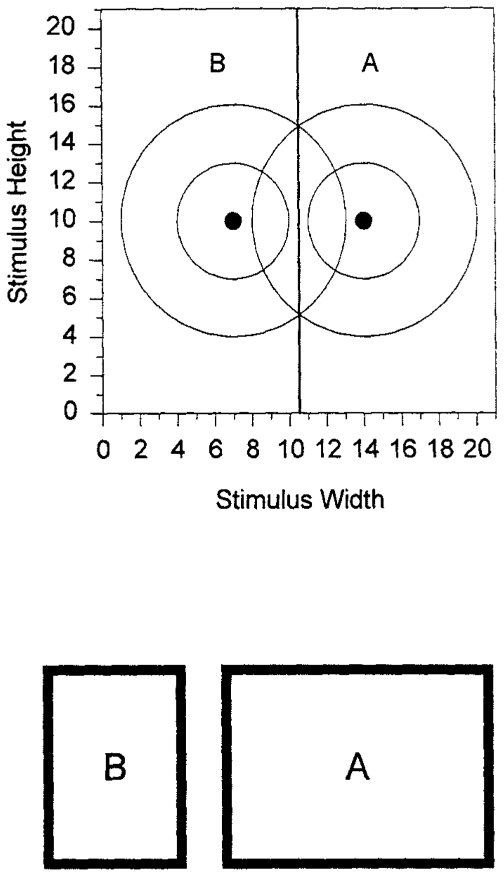


Figure 1.16. Herbranson, Fremouw & Shimp 1999, Stimuli distributions and rectangle examples for the RB task.

Notably the distributions in Herbranson et al.'s (1999) procedure overlap (see top panels of Figure 1.15 & 1.16), and thus the categories can be considered ill-defined. Because of this perfect performance is impossible, even if the subject adopts the optimal decision bound. This type of procedure and stimuli distribution was used by the authors with the primary goal of examining optimality in an ill-defined categorization situation.

The results from the divided-attention condition showed that pigeons were able to perform successfully in the task, and with an accuracy level that was close to optimal. Estimates of the decision bound were made for individual pigeons, and these were generally close to the ideal decision bound indicated by the diagonal line in the upper panel of Figure 1.15. Results from the selective-attention condition were similar, and showed that pigeons could also respond at a near-optimal level, consistent with decision bounds that were close to those shown in the upper panel of Figure 1.16. In a second experiment, Herbranson et al. showed that pigeons could respond accurately in a divided-attention condition in which the decision bound had a negative slope, and in a condition in which the optimal decision bound was nonlinear. Herbranson et al. concluded that pigeons could respond at near-optimal levels in all of the conditions they examined, and suggested that the randomization technique could be a useful methodology for studying naturalistic visual concepts in pigeons. Specifically, the randomization technique includes a number of features, such as large number of exemplars per category, and ill-defined or ‘fuzzy’ category boundaries. Herbranson et al. suggested that a theory or model that could account for pigeons’ performance in the randomization procedure might therefore yield insights into how pigeons (and other avians) learn naturalistic visual concepts.

However, Herbranson et al. (1999) noted one methodological aspect of their procedure that might have been problematic: the use of rectangles that varied in terms of height and width as stimuli. Previous research with humans by Krantz and Tversky (1975) had studied perception of rectangles. They had 17 subjects rate the degree or similarity or dissimilarity of 92 pairs of rectangles along a scale of 1-20. Results suggested that the dimensions of the rectangles (e.g. height, width, shape and area) were not independent and separable. Analysis of the data showed that area and shape dimensions provided a better account for those data than did width and height. Results suggested that subjects perceived the rectangles as skinny or fat shapes, rather than rectangles that could be described separately in terms of width and height. In the discussion the authors sum the results up well, “The data reject the hypothesis that area and shape contribute independently to the overall similarity between rectangles. They also show that the interaction between the dimensions is very orderly and readily interpretable” (Krantz & Tversky, 1975, p. 31). Krantz and Tversky’s results suggest that height and width are not separable dimensions. In this case, pigeons in Herbranson et al.’s (1999) study may not have been integrating information in the ‘divided attention’ task, that is, combining values of separate height and width dimensions prior to making a category decision (Massaro & Friedman, 1990). Thus, an important question is whether pigeons can respond accurately in such a task with stimuli that are comprised of fully-separable dimensions.

The research reported in this thesis attempts to address this issue by studying performance of pigeons in tasks similar to those used by Ashby, Maddox and colleagues and Herbranson et al. (1999), based on Gabor stimuli (Yao, Krolak, & Steele, 1995). Gabor stimuli can vary in terms of two separable dimensions, frequency and orientation. In addition, we used category distributions that did not overlap. This point of difference from Herbranson’s et al.’s (1999) work changes the criteria for optimal behavior in this type of task. In Herbranson et al., pigeons responded at levels that were close to the optimal decision bound, but the categories overlapped such that perfect performance was impossible. Thus we planned to revisit the question of whether pigeons' performances approached optimality in a task in which they could potentially reach 100% accuracy. .

Goals of the Present Research

Our first major question was to explore whether pigeons were capable of responding accurately in the RB (selective attention) and II (divided attention) tasks when stimuli were dimensionally-separable Gabor patches. We were particularly interested in the II task, because accurate performance with Gabor stimuli would indicate that pigeons were capable of information integration – that is, combining information from more than one dimension when making a category response (Massaro & Friedman, 1990). Model comparison analyses based on the GLC were planned to determine whether responding of pigeons reflected control by multiple or single dimensions. Experiments reported in Chapter 2 address these questions.

In addition, we wanted to compare performance of humans and pigeons in this categorization paradigm. There have been relatively few comparative studies of category learning by humans and non-humans that have used identical stimuli and procedures (for an exception, see Smith et al., 2004). In contrast to previous studies with humans by Maddox, Ashby and colleagues, which have generally used between-subjects designs in which participants completed the experiment in a single session, we planned to conduct our human research using a within-subjects design where the subjects each participated in all the conditions of the experiments in counterbalanced controlled order, conducted over multiple sessions. Experiments reported in Chapter 3 compare performance of pigeons and humans in the II task.

Optimality was also the focus of our investigations. Because Herbranson et al. reported optimal or near-optimal performance with pigeons using the same randomization technique with overlapping category distributions, we sought to determine whether performance would reach optimal levels when perfect accuracy was possible. We also examined the issue of optimality in our human vs. pigeon comparisons in order to look for species specific differences in performance.

Previous research has generally compared the overall accuracy of different groups or participants (or examined which of several models fit the data better). Thus in our study we planned to investigate responding more closely, particularly asymptotic performances after responding had been stabilized. Our major question here was to explore whether the GLC provided an adequate account of responding by pigeons and humans in the II tasks. To anticipate, because the GLC was found not to provide an adequate account of the data, a new model for category learning is proposed in Chapter 4.

Chapter 2 - Categorization of Multidimensional Stimuli by Pigeons.

The goal of the experiments reported in Chapter 2 is to investigate pigeons' performance in a two-dimensional category learning task based on Ashby and Gott's (1988) randomization procedure with dimensionally-separable stimuli. In contrast to previous research with pigeons which has used rectangles that varied in height and width (Herbranson, Fremouw & Shimp, 1999) the stimuli used here were Gabor patches that varied in frequency and orientation. Experiment 1 compared performance in two conditions which varied in terms of whether accurate performance required control by both dimensions ("information integration; II) or by a single dimension ("rule based"; RB). Experiment 2 compared performance in two rule-based tasks in which either frequency or orientation was the relevant dimension. The major question was whether pigeons would be able to respond accurately in both tasks. If so, we planned to examine asymptotic performance closely to determine if the GLC provided a satisfactory account of the results.

INTRODUCTION

The ability to categorize stimuli in the natural environment is essential for survival. For example, a thirsty gazelle must choose a drinking area in a river that is free from crocodiles. Migratory birds must properly evaluate a wide variety of stimuli, including the pattern of the sun's travel in the sky, earth's magnetic fields, visual landmarks and olfactory information, to navigate successfully to a warmer climate (Walraff, 2005). The skipper of a fishing boat in the Bering sea must determine if the weather is safe based on detailed satellite images of cloud, temperature and atmospheric pressure. As these examples show, categorization is a vital skill for any organism to have.

A binary categorization task may be regarded as a conditional discrimination in which one of two responses is reinforced depending on whether a prior stimulus is a member of one class or another (Zentall, Galizio, & Critchfield, 2002). The study of categorization in nonhumans has typically used binary categorization tasks and adopted one of two strategies. One approach has been to examine the ability of subjects to categorize stimuli that are comparable in terms of complexity to those that might be encountered in the natural environment. For example, in a pioneering study, Herrnstein and Loveland (1964) showed that pigeons were able to respond differentially depending on whether or not a photograph project onto the front panel of an operant chamber contained people or not. Herrnstein, Loveland and Cable (1976) trained pigeons to discriminate pictures with or without trees, water, and a specific person. In all three experiments stimuli sets included images that were easy to discriminate with whole or large parts of a person, tree or water and also more difficult images with only small parts or even similar-looking components. Results showed that pigeons were able to classify novel exemplars from each category correctly. Herrnstein et al. concluded that it was unlikely the pigeons used a feature-based strategy to discriminate among the naturalistic categories, and illustrated their point by noting the difficulty of describing features that would reliably discriminate between pictures of a celery stalk and a tree (see Herrnstein et al., 1976, Figure 3). Other studies involving complex stimuli have shown that apes can distinguish real objects from their photographs (Davenport & Rogers, 1971), pigeons can distinguish between impressionist or cubist paintings, (Watanabe, Sakamoto, & Wakita, 1995) and California Sea lions have the ability to form equivalence classes with arbitrary non-natural figures (Kastak, Schusterman, & Kastak, 2001).

The second approach has examined organisms' ability to categorize stimuli that vary quantitatively along a single dimension. These studies arrange conditional discriminations in which the subject must make one of two responses depending on the value of a stimulus on a particular dimension. For example, a pigeons' response to the left key might be reinforced after a bright light has been presented on a center key, whereas a response to the right key might be reinforced after a dim light (e.g., Davison & McCarthy, 1989). Much of this research has attempted to test predictions of signal detection theory and related models for discrimination (Davison & Tustin, 1978; Davison & Nevin, 1999; White & Wixted, 1999; see Alsop, 2004 for review). These models provide an excellent account of results from experiments in which category structure is defined by a stimulus value on a single dimension.

Overall, the majority of research on categorization by nonhumans has used stimuli that are either highly complex, with category structures that may be impossible to distinguish in terms of specific features or dimensions, or else used stimuli that have varied along a single dimension. By contrast, there has been a relative lack of research in which categories are defined in terms of quantitative variation along more than one dimension. Such research would fill an important gap, because it would allow for the development of more complex and potentially more realistic models for category learning based on multidimensional stimuli. The goal of the present research is to investigate how pigeons categorize stimuli that vary quantitatively along two dimensions. In particular, we wanted to test whether pigeons were able to learn a category task that required a comparison of stimulus values from different dimensions – that is, whether pigeons are capable of information integration (Massaro & Friedman, 1990).

Research on category learning by humans has used tasks in which stimuli vary along multiple dimensions. For example, Ashby and Gott (1988) had participants categorize L-shaped line segments that differed in terms of the heights of the vertical and horizontal lines. To generate stimuli, Ashby and Gott used a randomization procedure in which categories were defined in terms of two bivariate normal distributions. The category dimensions were two dimensional (height and width), and the distributions that formed the categories overlapped such that perfect performance was impossible, but an optimal decision bound was formed by a diagonal line, with a slope of 1, that ran down the middle of each of the two distributions. Results showed that performance of the three subjects was close to optimal with an average accuracy of 83%. The authors described their results in terms of Ashby and Townsend's (1986) General Linear Classifier (GLC), which is effectively a generalization of signal detection theory. According to the GLC, subjects' performance in a two-dimensional categorization task is based on learning a decision bound which can be represented as a linear function in a two-dimensional stimulus space. Stimuli that lie above the decision bound are assigned to one category, whereas stimuli that lie below it are assigned to the other. The probability of an accurate decision is given by the distance of the stimulus from the decision bound. The GLC provides an important model for information integration – how subjects combine stimulus values from multiple dimensions in order to make a decision (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Maddox & Ashby, 2004; (Massaro & Friedman, 1990).

Herbranson, Fremouw and Shimp (1999) studied performance of pigeons in a task similar to that used by Ashby and Gott (1988). They trained pigeons to categorize rectangles displayed on a computer screen that varied in terms of height and width and were generated using two bivariate normal distributions that overlapped. In the 'divided attention' condition, accurate performance depended on both dimensions: Rectangles for which the height was greater than the width were likely to belong to Category A, whereas rectangles for which the width was greater than the height were likely to belong to Category B. In a second condition, 'Selective Attention', accurate performance depended on only one dimension. For example wide rectangles might belong to one category and narrow to the other, but the height of the rectangles was irrelevant. Results showed that the pigeons' performances in the tasks were close to optimal in both tasks. These results suggest that pigeons are capable of integrating information on the bases of 2 relevant dimensions.

Although Herbranson et al.'s (1999) results suggest that pigeons are capable of learning category structures that require values from different dimensions to be compared – that is, information integration (Massaro & Friedman, 1991) – such a conclusion requires the assumption that the height and width of the rectangles are fully separable dimensions. But as Herbranson et al. noted, this assumption may be problematic. In a study with humans, Krantz and Tversky (1975) found that similarity ratings for rectangular stimuli did not suggest that height and width were fully separable, and that subjects instead may have perceived differences between rectangles in terms of area and shape. In other words, rectangles which are taller than wide may have been perceived as 'skinny', whereas rectangles which are wider than tall may have been perceived as 'fat'. The implication is that accurate performance in Herbranson et al.' s Divided Attention condition may not require that the height and width of the rectangles be separately perceived, and thus their results do not constitute a strong test of information integration.

Subsequent research with humans has avoided this problem by using stimuli that have reliably separable and independent dimensions. For example, studies have used Gabor stimuli, which are computer-generated sinusoidal wave gratings that vary in terms of frequency and orientation modulated by a circular Gaussian filter (Yao, Krolak, & Steele, 1995). Using category structures similar to those employed by Herbranson et al. (1999) and Ashby and Gott (1988), research has shown that humans are capable of responding accurately in information integration tasks based on Gabor stimuli (Maddox, Ashby, & Bohil, 2003).

We describe two experiments which investigate whether pigeons can respond accurately in a two-dimensional categorization task using Gabor stimuli that varied in orientation and frequency. In Experiment 1, we explored performance in both an information integration task in which both orientation and frequency were relevant for discriminating the categories, and a rule-based task in which only frequency was relevant. In Experiment 2, we compared performance in two rule-based conditions in which either frequency or orientation was relevant. Both experiments used stimuli that were similar to previous research with humans (Maddox et al., 2003). Our primary goals were to determine whether pigeons were capable of information integration when the stimulus dimensions were fully separable, to what degree their performance was optimal, and whether the GLC could provide an adequate account of the results.

EXPERIMENT 1

In Experiment 1 we used stimuli and category structures that were based on previous research by Maddox and colleagues (Maddox, Ashby, & Bohil, 2003), and tested pigeons’ performance in both a ‘divided attention’ or ‘information integration (II) condition and a ‘selective attention’ or rule-based (RB) condition. In both conditions, Gabor stimuli varied in orientation and frequency. In the II condition, accurate performance required attention to both dimensions, whereas in the RB condition, categories were distinguished in terms of frequency but not orientation. Unlike Herbranson et al.'s study, perfect performance was possible in both conditions because the stimuli from the categories did not overlap. To test whether pigeons were capable of information integration, we compared fits of the GLC against unidimensional models to test whether performance showed control by one or two dimensions. Finally, we wanted to assess whether the GLC was an adequate model for performance in this task and thus planned regression analyses to test whether systematic deviations in the fits of the GLC to the data obtained.

METHOD

Subjects

Six pigeons, designated H2, H3, H4, H5, H7, and H8, participated as subjects and were maintained at 85% of free-feeding weight ± 15 g by post-session feedings. They were housed individually and allowed free access to water and grit, in a vivarium with a 12:12 hr light/dark cycle (lights on at 7:00 a.m.). All had no prior experimental histories.

Apparatus

Four operant chambers, 350 mm deep by 360 mm wide by 350 mm high, were used. One wall contained an aluminum response panel in which a VGA 6.4” (130 mm wide x 97 mm tall) LCD display set to 640 x 480 resolution was mounted The LCD display was located 165 mm from the side edge and 230 mm from the bottom floor to center of the screen. Overlaying the LCD screen was a glass panel mounted resistive touch screen of identical size to the screen with a 4096 x 4096 point array resolution. Screen responses were measured via a USB touch interface (Elo TouchSystems Inc). The displays with touch panels were purchased from Touch Screens Inc, part number MTF064D. There were two vertically-aligned response keys on each side of the screen, midway between the edge of the screen and the chamber wall. The keys were 25 mm in diameter, and could be illuminated with 5 color LED arrays. A force of approximately 0.10 N was necessary to operate each key, and produced an audible feedback click. Centered below the screen was a grain magazine with an aperture (60 mm by 50 mm) 40 mm above the floor. The magazine was illuminated when wheat was made available by a white LED. A houselight was centered above the LCD screen 10 mm from the top of the panel. Chambers were enclosed in a sound-attenuating box, and ventilation and white noise were provided by an attached fan. Event scheduling, data recording, and screen image display was controlled with an IBM®-compatible microcomputer. Chamber keys, grain magazine and all other hardware inputs and outputs were interfaced via a USB module with 24 bits of digital I/O purchased from Measurement Computing (part # USB-1024LS).

Stimuli

The stimuli for the categorization tasks were Gabor patches. Gabor patches are sine wave gratings modulated by a circular Gaussian filter, and vary in terms of frequency and orientation. Sample Gabor patches are shown in Figure 1&2.

Two sets of Gabor stimuli were produced to yield two different types of categorization tasks (Maddox, Ashby, & Bohil, 2003). Each of the two stimuli sets can be represented in a 2 dimensional space with orientation on the X axis and frequency on the Y axis. For the RB condition, the optimal decision bound was a horizontal line drawn through the scatter plot (shown in Figure 2.1), representing a criterial value, such that stimuli with frequencies less than the criterion were assigned to one category, while stimuli with frequencies greater than the criterion were assigned to the other category. The stimuli for the II condition were obtained by rotating the stimuli from the RB condition 45 degrees to the right. The decision bound, scatterplot and 11 Gabor patches from each of the two categories are shown in Figures 2.1& 2.2. Sample Gabor patches are also displayed in figures, which are 1/6 scale sized images from the actual images used in the sessions. The examples include the extreme values for each category (i.e., the stimuli in the lower left and upper right of the scatterplot) and also nine stimuli in between, spaced approximately equally. The example stimuli that are pictured correlate to the filled symbols in the plots for each category.

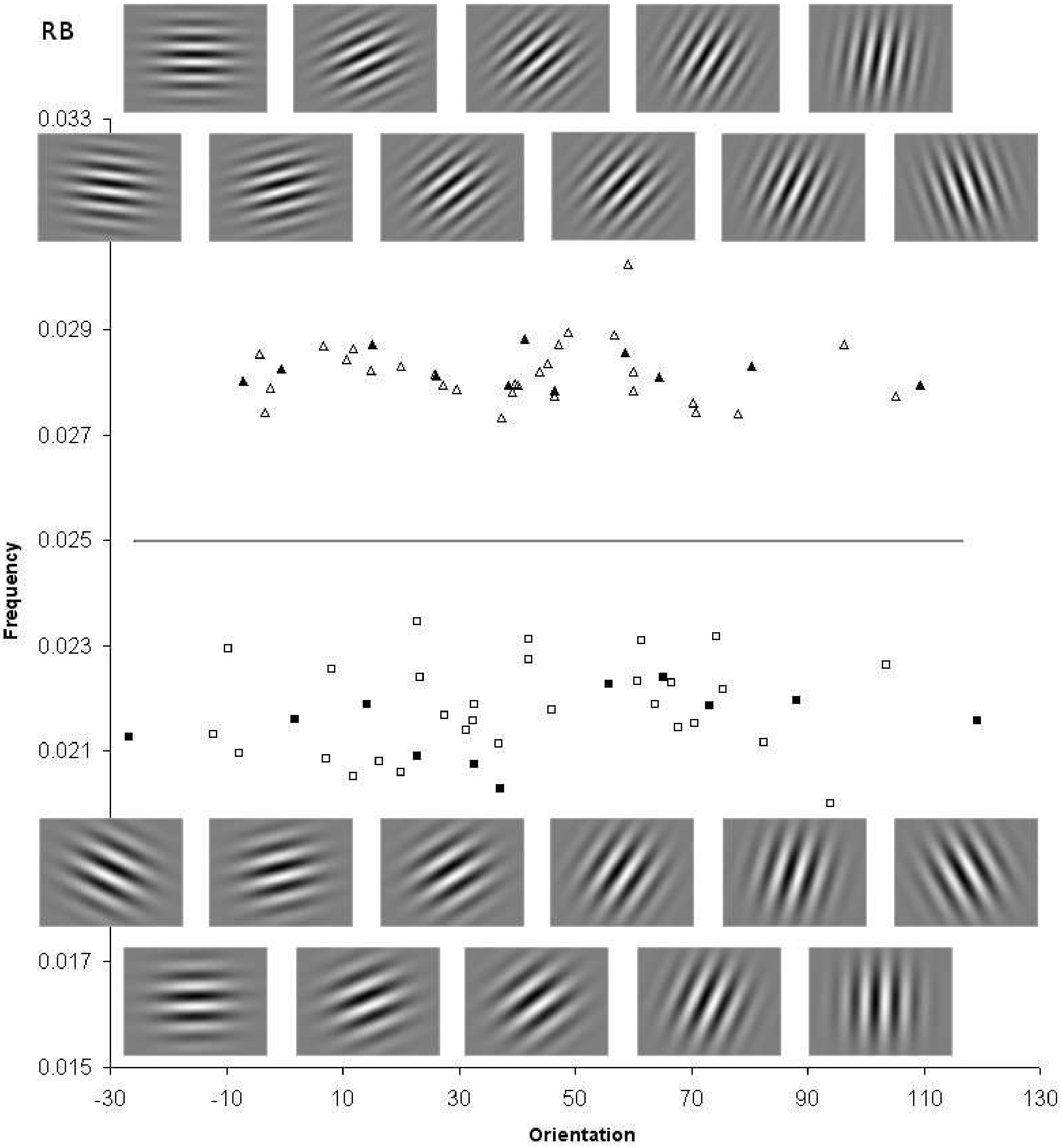


Figure 2.1. The Rule Based (RB) stimuli distribution as well as sample gabor images of each of the darkened stimuli. The X axis shows the orientation and the y axis the frequency. The solid line represents the optimal decision bound.

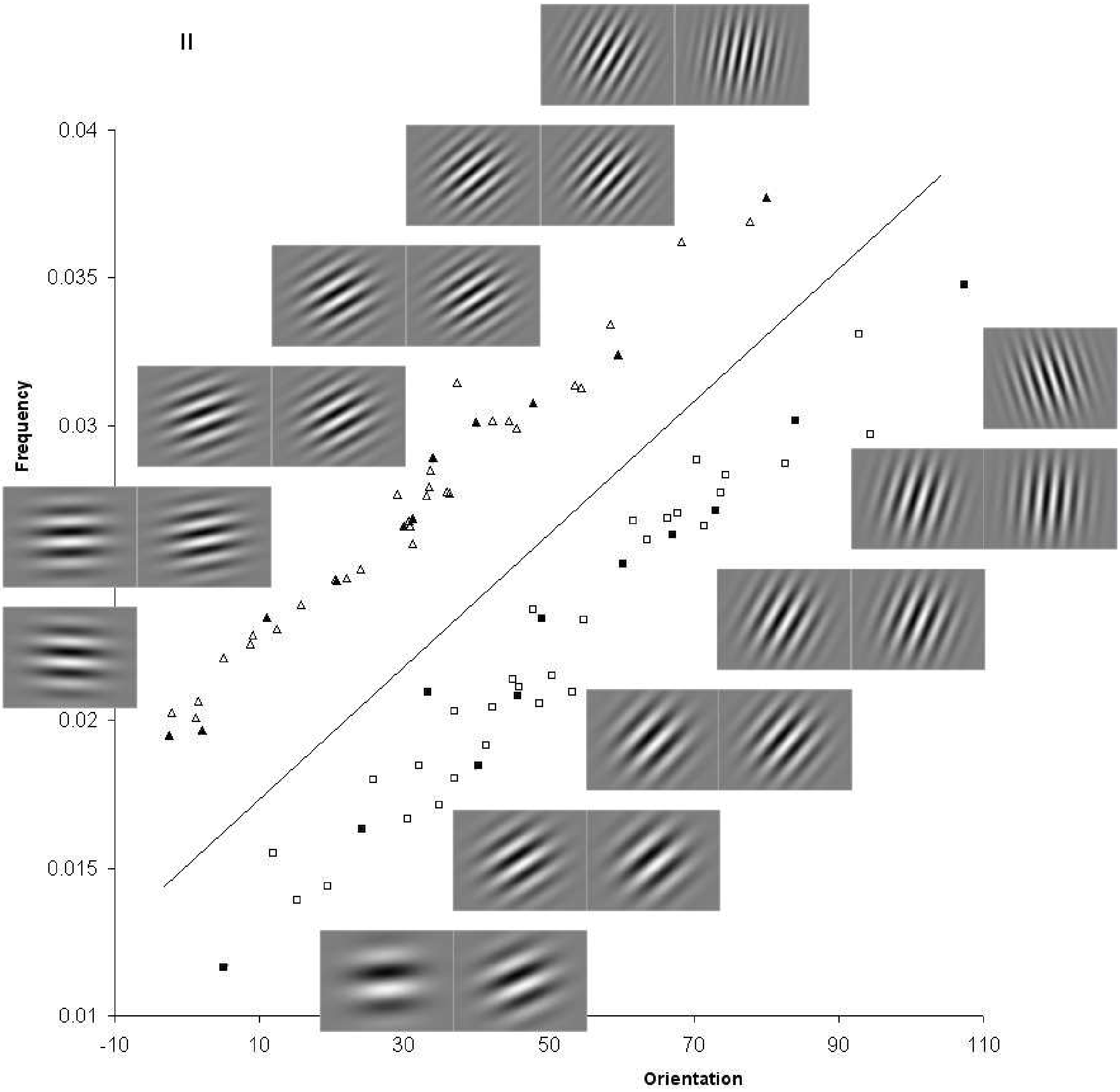


Figure 2.2. Set up the same as Figure 1 but displays the Information Integration (II) set of stimuli, examples and decision bound.

Means and standard deviations, as well as maximum and minimum values for the stimuli in each category for both the RB and II condition are shown in Table 2.1.

Stimuli were generated as follows: First, for the RB stimuli, random numbers were sampled from a bivariate normal distribution for each of the categories, A and B. Forty number pairs were sampled for each category, defining 40 stimuli in terms of frequency and orientation. The distribution parameters for each category were the same as Maddox et al. (2003), and ensured that the mean frequency values (Means = 260, 340, SDs = 11, 8.66, for category A and B respectively) were different for the categories whereas the mean orientation values (Mean = 125, SD = 94.8) were the same. The II stimuli were generated by rotating the RB stimuli by 45 degrees. After rotation, the stimuli were subjected to a linear transformation (5.98 was added to each frequency value and 245.81 added to each orientation) so that the grand means for frequency and orientation (i.e., the averages across both categories) were the same in the II and RB condition.

	Rule Based (RB)				Information Integration (II)			
	Category A		Category B		Category A		Category B	
	Orient	Freq	Orient	Freq	Orient	Freq	Orient	Freq
Min	-26.67	0.0200	-6.96	0.0273	5.15	0.0116	-2.43	0.0195
Max	119.26	0.0234	109.36	0.0302	107.35	0.0348	80.09	0.0377
Difference	145.93	0.0034	116.32	0.0029	102.20	0.0232	82.52	0.0182
SD	33.86	0.0008	29.65	0.0005	23.65	0.0055	21.13	0.0047
Mean	41.91	0.0218	41.38	0.0282	52.08	0.0227	31.21	0.0272

Table 2.1. The distribution parameters for each category in each condition. Orientation (Orient) values are in degrees and frequency values are in cycles per pixel.

For display on the LCD screens (640 x 480 resolution), the values were converted to cycles per pixel with the following equation where x equals the randomly generated frequency values: (X/50+.25)/250. The result was a frequency value in cycles per pixel that was maximized for display on a 640 x 480 screen. For the orientation dimension, numbers were converted to degrees from horizontal by first multiplying by π/500, then to degrees from horizontal by multiplying the radian value by 180 then dividing by π.

Gabor stimuli were generated in real-time using custom software written at the University of Canterbury. The algorithm used was based on the Gabor Filter, (Yao, Krolak, & Steele, 1995), and was integrated into a C++ program that displayed the images based on a pre-determined CSV file listing of frequency (cycles per pixel) and orientation (degrees).

Procedure

Because subjects were experimentally naïve, they were first shaped to peck yellow circles displayed in the center of the touch screen. They were then trained to peck the two lower right and left side keys using a modified autoshaping procedure. When subjects responded consistently both to the touch screen and keys, training began in the first condition. Sessions occurred daily and at the same time (1100h) with few exceptions. All sessions consisted of 90 trials and sessions were run until stability was achieved in each phase of each condition.

The sequence of events on experimental trials was as follows. After a 9-s inter-trial interval (ITI) during which the chamber was dark, the houselight was illuminated. One second later, the trial began with the display of a Gabor image on the touch screen. The image was maximum possible size that could be shown (640 x 480 pixels) and measured approximately 95 cm high by 125 cm wide. After pigeons had made five responses to the image the screen was darkened and the two lower keys were illuminated (e.g., left key red, right key green), signaling the choice phase. A single response to the correct key produced 3-s access to grain. During reinforcement, all illumination in the chamber was extinguished except for the feeder light. If the response was incorrect, the houselight flashed off and on for 10 s (1 s off, 1 s on), and the trial was repeated with the same Gabor stimulus. After five responses had been made to the screen, only the correct side key was lit and a single response produced 1.5-s access to grain.

Pigeons were exposed to the RB and II conditions in counterbalanced order, followed by a replication of the II condition. The replication was completed after the pigeons had participated in an unrelated experiment involving different Gabor stimuli. Training continued in each condition until a visual stability criterion was reached. In the first condition, extended training was given because we wanted to assess the long-term stability of responding given the novel nature of the procedure.

RB & II Condition Order							II Replication Condition		
Pigeon	Stimuli	Cat A Key	# Sessions	Stimuli	Cat A Key	# Sessions	Stimuli	Cat A Key	# Sessions
H2	RB	Left Red	89	II	Left Red	43	II	Left Red	34
H3	RB	Left Red	94	II	Right Green	33	II	Left Red	35
H4	RB	Left Red	65	II	Left Red	67	II	Left Red	34
H5	II	Left Red	70	RB	Right Green	35	II	Left Red	32
H7	II	Left Red	59	RB	Left Red	41	II	Left Red	32
H8	II	Left Red	69	RB	Right Green	35	II	Left Red	31

Table 2.2. The condition order for each of the pigeons. The stimuli, key counter balancing and number of sessions are also displayed.

The keys assigned to the categories, correct key location and color were counterbalanced across birds and are listed in Table 2.2, along with the order of conditions and number of sessions of training.

RESULTS

Figure 2.3 shows the percentage of correct choice responses for all subjects across the three conditions (II, RB, and II replication) in the experiment. The dashed line indicates chance 50% responding. We continued to run the sessions in the first condition well beyond asymptotic performance both due to the novelty of the procedure, and our desire to insure that we had indeed reached a stable level of performance. All pigeons acquired each of the tasks successfully, in terms of responding at greater than chance accuracy, although differences between the birds' performances are evident. Accuracy was relatively low for Pigeon H3 in the II and RB conditions, but increased in the II replication condition. For the other pigeons, accuracy tended to stabilize at levels between 75% and 85% in each of the conditions. Because perfect performance was possible, this indicates that all of the pigeons' performances fell short of optimality.

Average accuracies from the last 10 sessions of each condition are reported in Table 3 for each pigeon, as well an overall average. The averages were 83% (SD = .037), 82% (SD = .049) and 88% (SD = .032) correct for the II, RB, and II replication conditions, respectively. A repeated-measures analysis of variance (ANOVA) found that the effect of condition was not significant, $F(2,10) = 2.19$, $p > .15$.

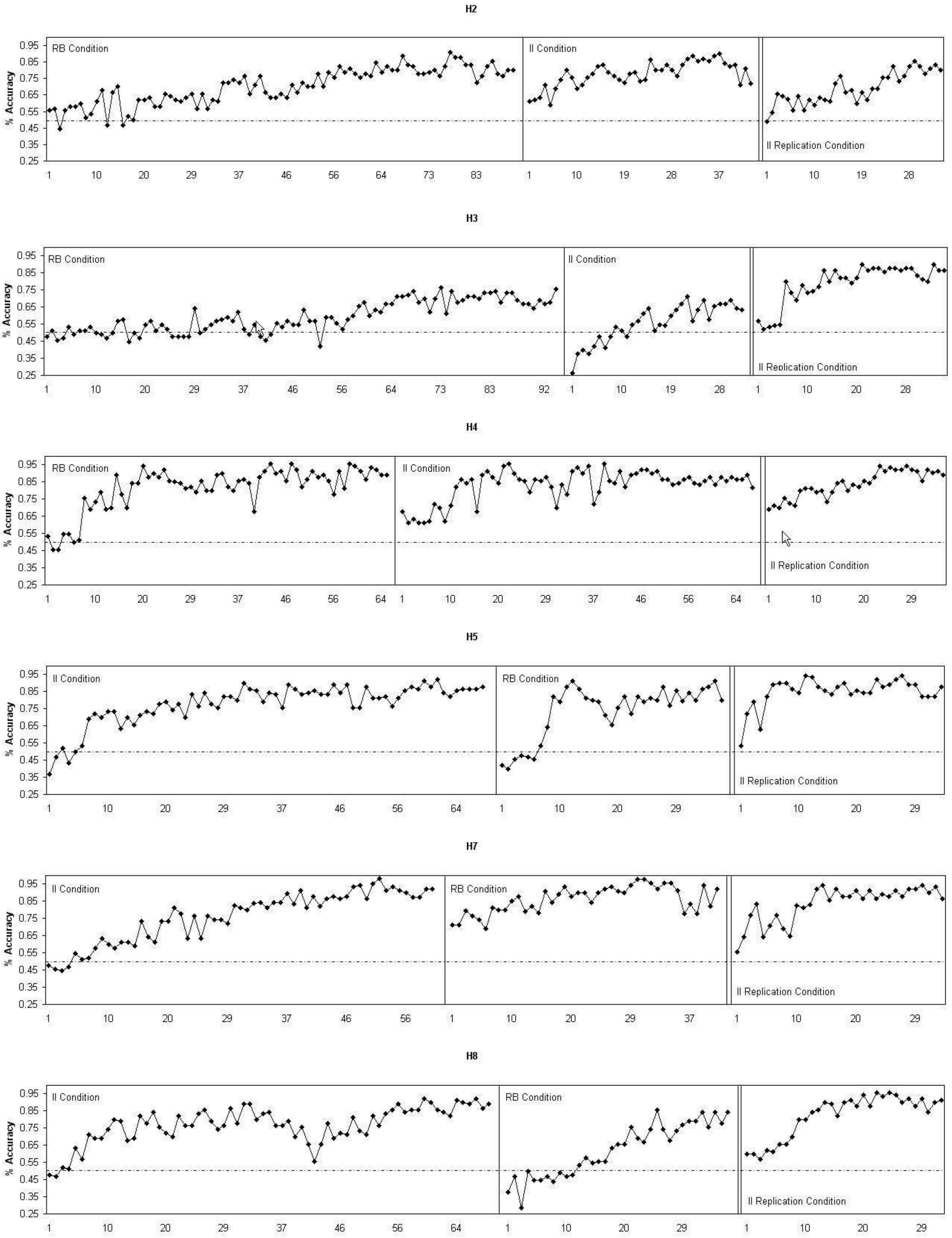


Figure 2.3. The mean accuracy for each of the sessions for each pigeon, x axis indicates the session # and y axis the % accuracy. II, RB and II replication condition are also labeled in the figure legend.

This suggests that there were no systematic differences in asymptotic accuracy between the conditions.

To investigate whether different amounts of training were necessary for the pigeons to acquire the II and RB tasks, we defined a post-hoc acquisition criterion of an average of 75% accuracy across the last three sessions, and then determined how many sessions were required to reach this criterion, for each pigeon and condition. Table 2.3 shows the results. Pigeon H3 never reached the 75% criteria in the first two conditions, but did so after 11 sessions in the II Replication condition, which was below the average for the rest of the pigeons. Averaged across pigeons (omitting H3's data from the first two conditions), 14.40, 22.20 and 12.17 sessions were required to reach 75% accuracy in the II, RB, and II replication conditions, respectively. To compare sessions to criterion across conditions, we conducted a repeated-measures ANOVA (omitting the data from H3). The effect of condition was not significant, $F(2,8) = 0.78$, $p > .40$. This indicates that there were no systematic differences in rate of acquisition across conditions.

	II		RB		II Replication	
Pigeon	% Accuracy	Sessions to 75%	% Accuracy	Sessions to 75%	% Accuracy	Sessions to 75%
H2	0.83	9	0.80	59	0.80	26
H3	0.64	-	0.69	-	0.86	11
H4	0.86	12	0.90	14	0.91	9
H5	0.87	19	0.84	10	0.88	6
H7	0.92	20	0.88	4	0.90	11
H8	0.88	12	0.78	24	0.91	10
Overall	0.83	14.40	0.82	22.20	0.88	12.17

Table 2.3. Mean accuracy for each pigeon's condition and the number of sessions to the stability criteria. Note that pigeon H3 never reached stability in the II or RB condition.

Model Analyses

In order to determine whether pigeons' responding in the RB and II conditions indicated control by a single stimulus dimension or by multiple dimensions, we conducted a series of modeling analyses. In these analyses, we compared fits of unidimensional and multidimensional models of category learning (Ashby, 1992) to individual-subject data from the last 10 sessions (900 trials) of each condition. We first provide a brief description of each of the models.

Multidimensional model (General Linear Classifier)

According to the General Linear Classifier (GLC), which is one of a family of models known as General Recognition Theory (GRT; Ashby, 1989; Ashby & Gott, 1988; Ashby & Townshend, 1986), stimuli are represented in a two-dimensional perceptual space, similar to Figures 1 and 2. The subject learns to associate different regions of the perceptual space with different responses. The two regions in the perceptual space are defined by a linear 'decision bound'. When a stimulus is presented on a given trial, the distance of the stimulus from the decision bound determines the probability of a choice response. Specifically, the decision bound is defined as:

$$\delta X + \gamma Y + e = 0 \qquad (1)$$

Where X and Y are orientation and frequency, respectively, and δ, γ, and e are constants. When a stimulus X_0, Y_0 is presented on a trial, the distance of the stimulus from the decision bound is given by:

$$h(X_0, Y_0) = \frac{\delta X_0 + \gamma Y_0 + e}{\sqrt{\delta^2 + \gamma^2}} \tag{2}$$

For $h = 0$, the probability of responding category A, $p(\mathrm{A}) = 0.50$. For $h > 0$, $p(\mathrm{A}) > 0.50$ and for $h < 0$, $p(\mathrm{A}) < 0.50$. Specifically, $p(\mathrm{A})$ is given as the cumulative normal distribution function (Φ) evaluated at $h(X_0,Y_0)$:

$$p(A) = \Phi\left(\frac{h(X_0, Y_0)}{\sqrt{\sigma_h^2 + \sigma_c^2}}\right). \tag{3}$$

The denominator of Equation 3 represents the noise or error variance in the model, and includes terms for both perceptual (σ_c^2) and criterial variance (σ_c^2). Although other models within GRT can distinguish between perceptual and criterial variance (see Ashby, 1992), for the GLC only a single parameter, σ, is estimated which represents combined perceptual and criterial variance. Effectively, the GLC represents a generalization of signal detection theory to the two-dimensional case (Ashby & Townshend, 1986).

In applying the GLC to data from the present experiment, three parameters must be estimated: the slope and intercept of the decision bound, and the noise parameter, σ. Note that the slope and intercept are defined as $-\delta/\gamma$ and $-e/\gamma$, respectively.

Unidimensional models

Two unidimensional models were also considered. According to the UNI-O model, subjects respond on the basis of orientation, but variation in frequency has no effect. The UNI-F model is similar except that decisions are based entirely on frequency. These models could be considered as special cases of the GLC in which the decision bound is represented as a straight horizontal line (UNI-F) or straight vertical line (UNI-O) in Figures 2.1 and 2.2. Both models have two parameters: a critical value on the particular dimension ($Xcrit$) and a noise parameter, σ. For stimulus X presented on a given trial, the probability of responding category A is defined as

$$p(A) = \Phi\left(\frac{X - Xcrit}{\sigma}\right) \qquad (3)$$

Parameter estimation

Maximum likelihood estimation was used to obtain parameters for the GLC and unidimensional models for individual-subject data. Specifically, parameter values that minimized the negative log-likelihood function were obtained through a two-step process. First, a simulated annealing algorithm (Goffe, Ferrier, & Rogers, 1994) was used to estimate a local minimum, and then parameter estimates were refined using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method (Avriel, 2003). Initial parameter values were randomly determined. Model predictions and optimization procedures were implemented in a computer program using routines in the open-source TPMATH library and compiled with Free Pascal version 2.0.2 (retrieved on 27 August 2006 from http://www.unilim.fr/pages_perso/jean.debord/tpmath/tpmath.htm and http://www.freepascal.org, respectively). Repeated simulations showed that parameter estimates were stable for all subjects and conditions and did not depend on initial values.

Model Comparison

Model fits for all subjects and conditions were evaluated using the Akaike Information Criterion (AIC; Akaike, 1974). The AIC is a model comparison statistic and defined as

$$AIC = -2\ln L + 2v\,, \tag{4}$$

where L is the likelihood function and v is the number of parameters estimated. AIC can be used to compare the adequacy of fits for different models applied to the same data: The model with the lowest AIC value has the best fit. For each data set, Table 4 indicates the best-fitting model by displaying the lowest AIC value in boldface. Table 4 also shows the variance accounted for (VAC) by each model.

Results in Table 4 show that for both the original II and replication conditions, the GLC model had the lowest AIC value in 11 out of 12 cases (the exception being H3, original II condition, for which the UNI-O model had the lowest AIC, suggesting that responding was controlled exclusively by orientation). Averaged across subjects, the variance accounted for by the GLC in the original II and replication conditions was 0.89 and 0.88 respectively. This confirms that pigeons’ choice responding was determined by

	GLC		UniF		UniO	
	AIC	VAC	AIC	VAC	AIC	VAC
H2CII	**73.67**	0.89	942.99	-1.45	110.92	0.07
H2CRB	81.28	0.83	**77.34**	0.82	112.00	0.00
H2CII	**84.46**	0.79	107.94	0.08	102.82	0.21
H3CII	84.08	0.83	104.73	0.24	**80.33**	0.83
H3CRB	105.15	0.58	**101.94**	0.55	114.18	0.03
H3CII	**78.45**	0.87	112.17	0.03	99.00	0.34
H4CII	**66.62**	0.87	110.06	0.09	101.45	0.25
H4CRB	55.90	0.95	**51.91**	0.95	114.25	0.00
H4CII	**64.74**	0.90	105.07	0.18	106.71	0.15
H5CII	**66.18**	0.92	108.58	0.12	103.35	0.22
H5CRB	76.13	0.90	**72.17**	0.90	114.76	0.00
H5CII	**61.55**	0.90	106.09	0.15	104.33	0.18
H7CII	**54.55**	0.96	2283.31	-1.29	101.09	0.23
H7CRB	51.96	0.96	**47.98**	0.96	113.13	0.00
H7CII	**56.27**	0.90	109.79	0.08	97.99	0.27
H8CII	**60.86**	0.85	109.53	0.06	97.10	0.25
H8CRB	80.46	0.88	**83.94**	0.71	107.09	0.14
H8CII	**56.28**	0.93	107.15	0.13	101.74	0.22

Table 2.4. Akaike information criteria (AIC) and variance accounted for (VAC) values are displayed for each of the pigeons in each condition for each of the 3 models tested. General linear classifier (GLC, unidimensional frequency (UniF) and unidimensional orientation (UniO) models were tested.

	GLC			UniF		UniO	
	Slope	Intercept	Noise	Mean	Sigma	Mean	Sigma
H2CII	0.011	0.016	0.004	-0.024	0.010	0.962	1.713
H2CRB	0.000	0.024	0.004	0.024	0.003	91.207	371.505
H2CII	0.014	0.017	0.006	0.031	0.022	0.418	0.984
H3CII	0.139	-0.073	0.054	0.025	0.011	0.702	0.413
H3CRB	0.002	0.024	0.006	0.025	0.006	0.525	4.857
H3CII	0.017	0.014	0.005	0.030	0.031	0.590	0.716
H4CII	0.015	0.014	0.004	0.024	0.018	0.772	0.779
H4CRB	0.000	0.024	0.002	0.024	0.002	96.755	784.682
H4CII	0.012	0.016	0.004	0.026	0.012	0.668	1.027
H5CII	0.014	0.015	0.004	0.026	0.015	0.693	0.833
H5CRB	0.000	0.025	0.003	0.025	0.003	0.029	17.092
H5CII	0.013	0.015	0.004	0.024	0.013	0.816	0.906
H7CII	0.014	0.015	0.003	0.727	0.085	0.669	0.757
H7CRB	0.000	0.024	0.002	0.024	0.002	91.954	476.486
H7CII	0.015	0.014	0.004	0.026	0.017	0.671	0.668
H8CII	0.015	0.015	0.004	0.028	0.019	0.583	0.656
H8CRB	-0.003	0.026	0.004	0.024	0.004	0.375	1.565
H8CII	0.014	0.014	0.004	0.024	0.014	0.794	0.790

Table 2.5. Model parameters for each of the models tested indicated by slope, intercept and noise.

both stimulus orientation and frequency values in the II task (Figure 2.2). This suggests that the pigeons were able successfully to integrate information from two stimulus dimensions in a category-learning task with Gabor stimuli.

Table 4 also shows that for the RB condition, the UNI-F model had the lowest AIC value for all subjects. Across subjects, the average variance accounted for by the UNI-F model was 0.81. This shows that when frequency was the only relevant dimension, it acquired primary control over choice responding.

Parameter values for each model are listed in Table 2.5. Overall, GLC parameter values were reasonably consistent across subjects and replications of the II condition. Estimates of noise parameters were also similar across replications of the II condition, and in the RB condition.

Detailed Analyses of Asymptotic Performance

For a more detailed investigation of how asymptotic performance was related to stimulus characteristics, we plotted the probability of a choice response for category A as a function of orientation for all subjects and conditions. Figures 2.4 through 2.6 show the resulting scatterplots for the II, RB, and II replication conditions, respectively. Category A stimuli are indicated by unfilled triangles, and Category B stimuli are indicated by filled squares. The overall accuracy (percentage correct) is also displayed in the upper right corner of each panel. Note that results are shown as a function of orientation only in these figures for sake of economy. Because orientation and frequency were positively correlated in the II conditions, results would look similar if plotted as a function of frequency. For the RB condition, frequency was the relevant dimension and its control has already been established through the model fits; because orientation was irrelevant, plotting the data as a function of orientation should reveal no systematic pattern.

For the original II and replication conditions (Figures 2.4 and 2.6), a systematic pattern of responding was found for category A. With the exception of H3, original II condition, *p*(A) for category A stimuli was an inverted-U shaped function of orientation for all subjects. Because results for H3, original II condition were also exceptional in that the UNI-O model provided the best fit, this means that the inverted U-shaped pattern was obtained in every case for which the GLC provided the best account of the data. The implication is that for category A, accuracy was greater for stimuli that were in the middle of the range of orientations, compared to stimuli with orientations that were near the ends of the range. More variable results were obtained for Category B stimuli. In the majority of cases, there was no systematic relationship between choice responding and orientation. However, for H2 in the original II condition, and H4 and H5 in the replication, *p*(A) tended to increase with increases in orientation. A similar pattern was obtained for H3, original II condition, consistent with the control by orientation obtained in the model fits (Table 2.4).

Figure 5 shows the results for the RB condition. For category A stimuli, a similar inverted-U shaped pattern was obtained for *p*(A) as a function of orientation for H2, H3, H4 and H7, although this pattern was less extreme than in the II conditions. For category B stimuli, results were variable: There was some indication of an inverted-U shaped pattern for H2 and H3, and *p*(A) increased with orientation for H8.

Results for the other pigeons showed no clear trend.

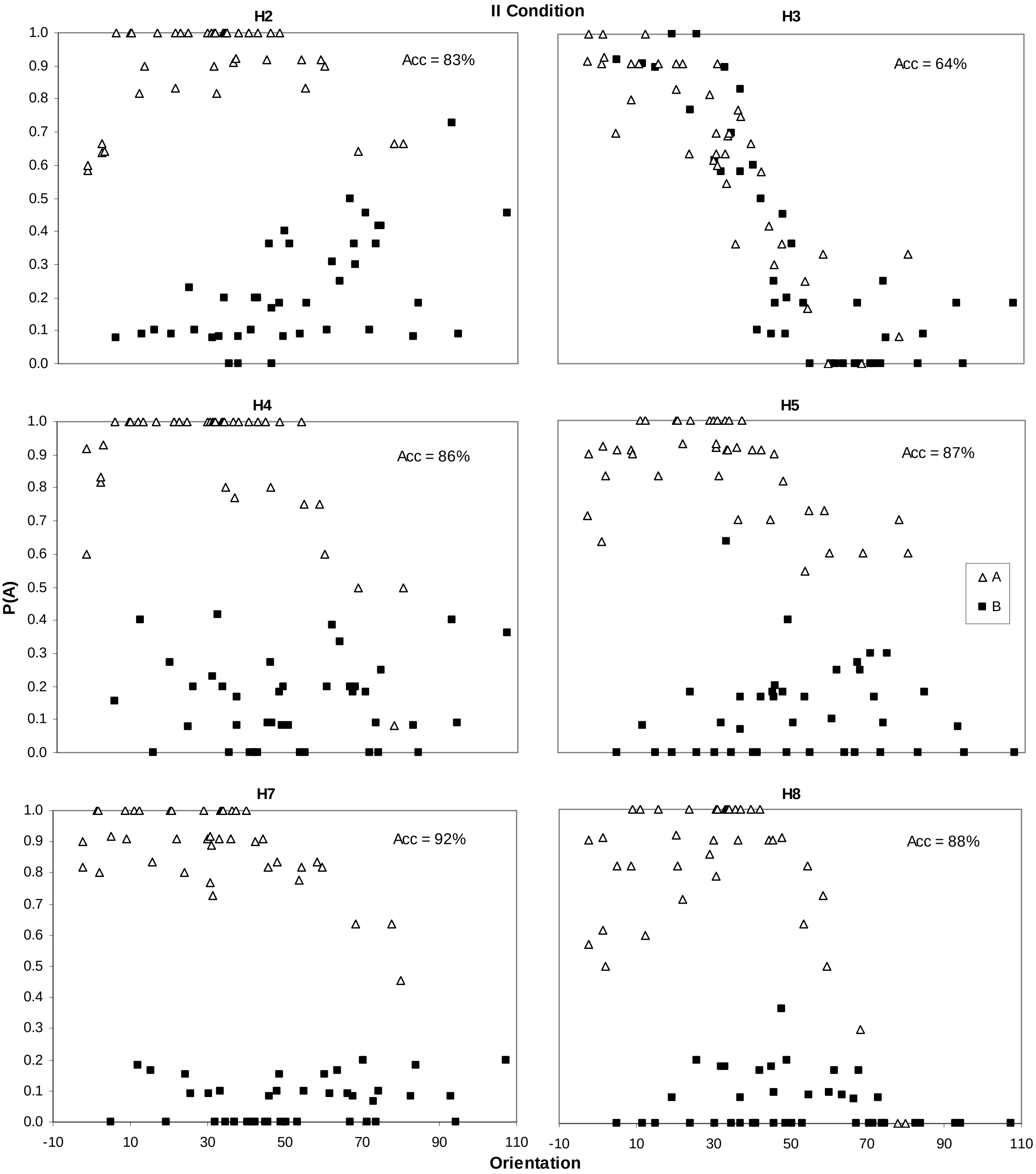


Figure 2.4. Probability of a category A response (P(A) in the y axis plotted against orientation in the x axis. The II condition is displayed for all 6 pigeons with category a represented by the unfilled triangles and b the filled squares.

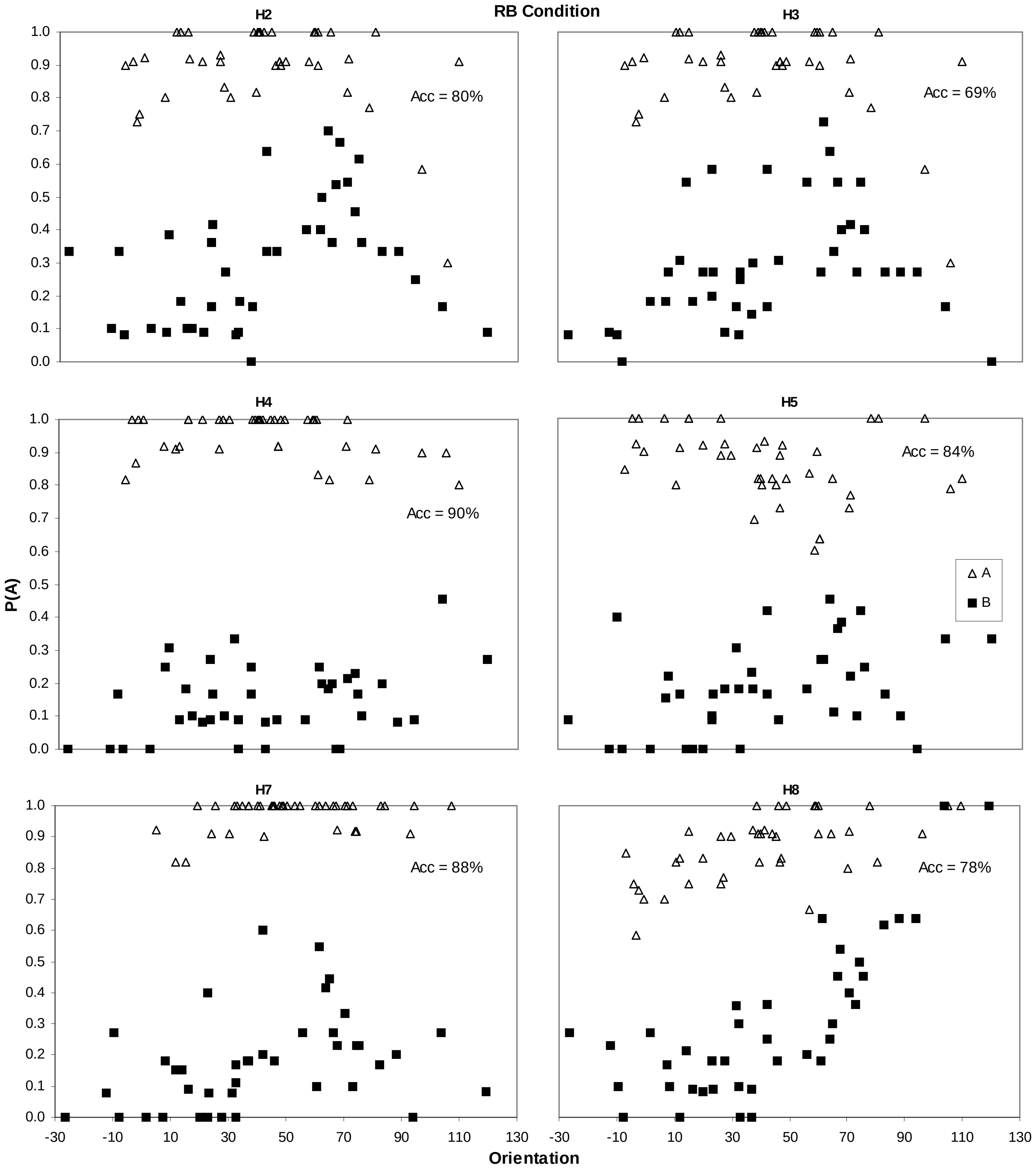


Figure 2.5. Same format as figure 2.4 but for the rule based (RB) condition.

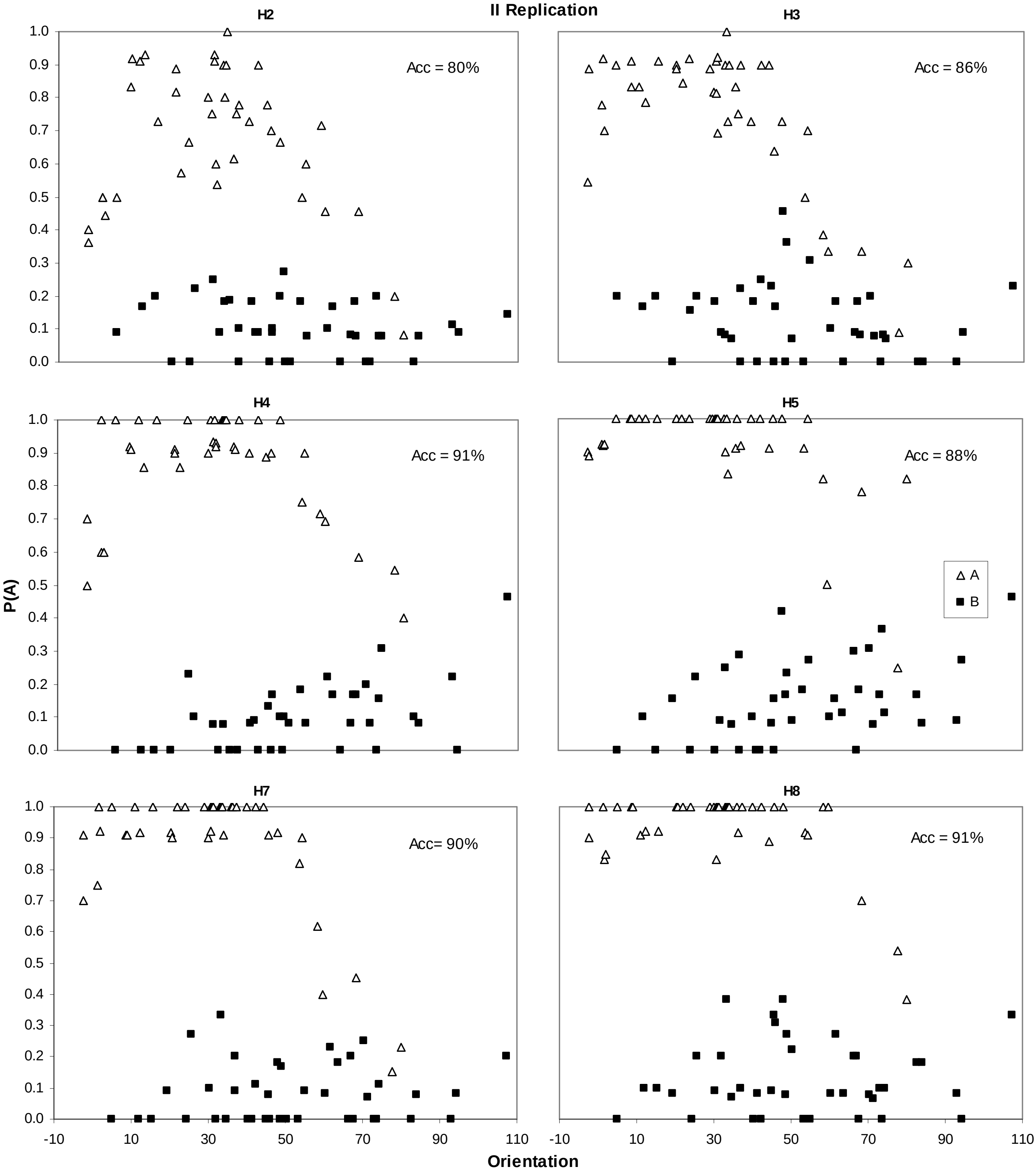


Figure 2.6. Information integration replication (II Replication) condition.

Overall, results in Figures 2.4 through 2.6 show that performance varied systematically as a function of orientation in both II and RB conditions. Results were most consistent across subjects for category A responding in the II conditions, for which in every case that accuracy levels were substantial overall (>75%) and the GLC was the best-fitting model, indicating that responding was controlled by both dimensions, accuracy was highest for orientations in the middle of the range (~45 degrees) and decreased as orientation tended to either extreme (0 or 90 degrees). It is important to note that the systematic patterns that were observed were all non-optimal, in the sense that optimal responding would have shown no within-category trend as a function of orientation, and the observed patterns were associated with decreased accuracy.

The systematic patterns in Figures 2.4 through 2.6 – especially the inverted-U shaped functions in the II conditions – represent a possible difficulty for the GLC as a model of category learning. Because the stimuli in the II condition were approximately equidistant from the optimal linear decision bound (see Figure 2.3), there would appear to be no way that the GLC could predict that performance should vary systematically as a function of orientation. Thus, to assess the adequacy of the GLC, we compared predicted and obtained values for GLC fits to individual data, and obtained standardized residuals for these fits.

Regression analyses

In order to test whether there were systematic deviations in the GLC residuals that might correspond to the patterns noted above in Figures 2.4 through 2.6, we conducted a series of polynomial regressions. Specifically, we used the orientation and the square of the orientation in a multiple regression to predict the standardized residuals. The orientation values were centered prior to squaring to avoid problems with multicollinearity. This analysis allows us to test the significance of both linear and quadratic relationships in the residuals. Regressions were performed for individual data for all conditions, as well as for the group mean data.

	Category A			Category B		
II	β Linear	β Quadratic	R^2	β Linear	β Quadratic	R^2
H2	0.09	-0.85 ***	0.69 ***	0.34	0.08	0.07
H3	0.4	0.02	0	-0.21	0.4 **	0.19 *
H4	-0.14	-0.77 ***	0.67 ***	0.35 *	0.30 *	0.24 **
H5	-0.14	-0.57 ***	0.39 ***	0.27	-0.36 *	0.17 *
H7	-0.23	-0.60 ***	0.49 ***	0.36 *	0.27	0.23 **
H8	0.11	-0.88 ***	0.73 ***	0.26	-0.46 **	0.24 **
Mean	-0.07	-0.92 ***	0.89 ***	0.48 **	0.22	0.31 **
RB						
H2	0.3	-0.24	0.13 *	-0.16	-0.55 ***	0.38 ***
H3	0.57 ***	-0.44 ***	0.46 ***	-0.62 ***	-0.24 *	0.51 ***
H4	0.15	0.16	0.05	-0.18	-0.49 **	0.31 **
H5	0.3	0.03	0.1	-0.46 **	0.35 *	0.25 **
H7	0.29	-0.28	0.15 *	-0.47 **	-0.26	0.36 ***
H8	0.09	0.67 ***	0.48 ***	-0.23	0.03	0.05
Mean	0.51 **	0.26	0.26 **	-0.65 ***	-0.34 **	0.65 ***
II Rep						
H2	0.2	-0.84 ***	0.64 ***	0.06	0.09	0.01
H3	-0.1	-0.73 ***	0.58 ***	0.31	-0.07	0.1
H4	-0.04	-0.84 ***	0.72 ***	0.39 **	0.21	0.22 **
H5	-0.30 *	-0.56 ***	0.50 ***	0.49 **	0.12	0.27 **
H7	-0.15	-0.83 ***	0.78 ***	0.52 ***	-0.13	0.26 **
H8	-0.12	-0.75 ***	0.62 ***	0.34 *	-0.03	0.11
Mean	-0.09	-0.90 ***	0.89 ***	0.58 ***	0.05	0.34 ***

Table 2.6. The left most column shows the label for each pigeon followed by the mean and the bold text divides the conditions, II, RB and II Replication. Category A is on the right set of results columns and category B on the left. Each category reports the Beta Weights (β) for the linear coefficient, β for the quadratic coefficients and the R^2 values. Significant regression results for β Linear or β Quadratic are indicated with an asterisk (*) for 0.1, 0.05, 0.01 significance levels using *, ** and *** respectively.

Results of the polynomial regressions are shown in Table 2.6. For both the II and II replication conditions, the quadratic coefficient for Category A residuals was negative and statistically significant for each subject and the group mean data with the exception of H3, original II condition. This means that a significant negative quadratic trend was obtained in the GLC residuals for each case in which the GLC provided the best fit to the data. This confirms that the GLC is unable to account for the inverted-U shaped pattern evident in Figures 4 and 6. For Category B residuals in the II conditions, linear coefficients were positive for all subjects with the exception of H3, original II condition. The positive linear coefficient was significant for H4 and H7 and the group mean data for the original II condition, and for H4, H5, H7, H8, and the group mean data in the II replication condition. This suggests that GLC predictions for Category B also showed systematic deviations from obtained values, with residuals tended to increase linearly with predicted values.

For results from the RB category, linear coefficients for Category A residuals were positive in all cases, and significant for H3 and the group mean data. For Category B residuals, linear coefficients were negative in all cases, and were significant for H3, H5, H7 and the group mean data. This suggests that the predictions of the GLC for the RB condition also showed systematic deviations from the data.

Figure 2.7 provides a summary of the residual analyses based on the group-mean data. The left panels show the obtained data for Category A and B (unfilled triangles and filled squares, respectively) and GLC predictions (x’s and +’s, respectively), whereas the right panels show the standardized residuals for Category A and B (unfilled triangles and filled squares, respectively). Results for the II, RB and II replication conditions are shown in the upper, middle, and lower row, respectively.

Figure 2.7 clarifies how the GLC has failed to describe systematic features in the current data. The inverted-U shaped pattern that is evident for Category A in the II conditions produces a sharp decrease in p(A) for high values of orientation, to levels below 0.50.

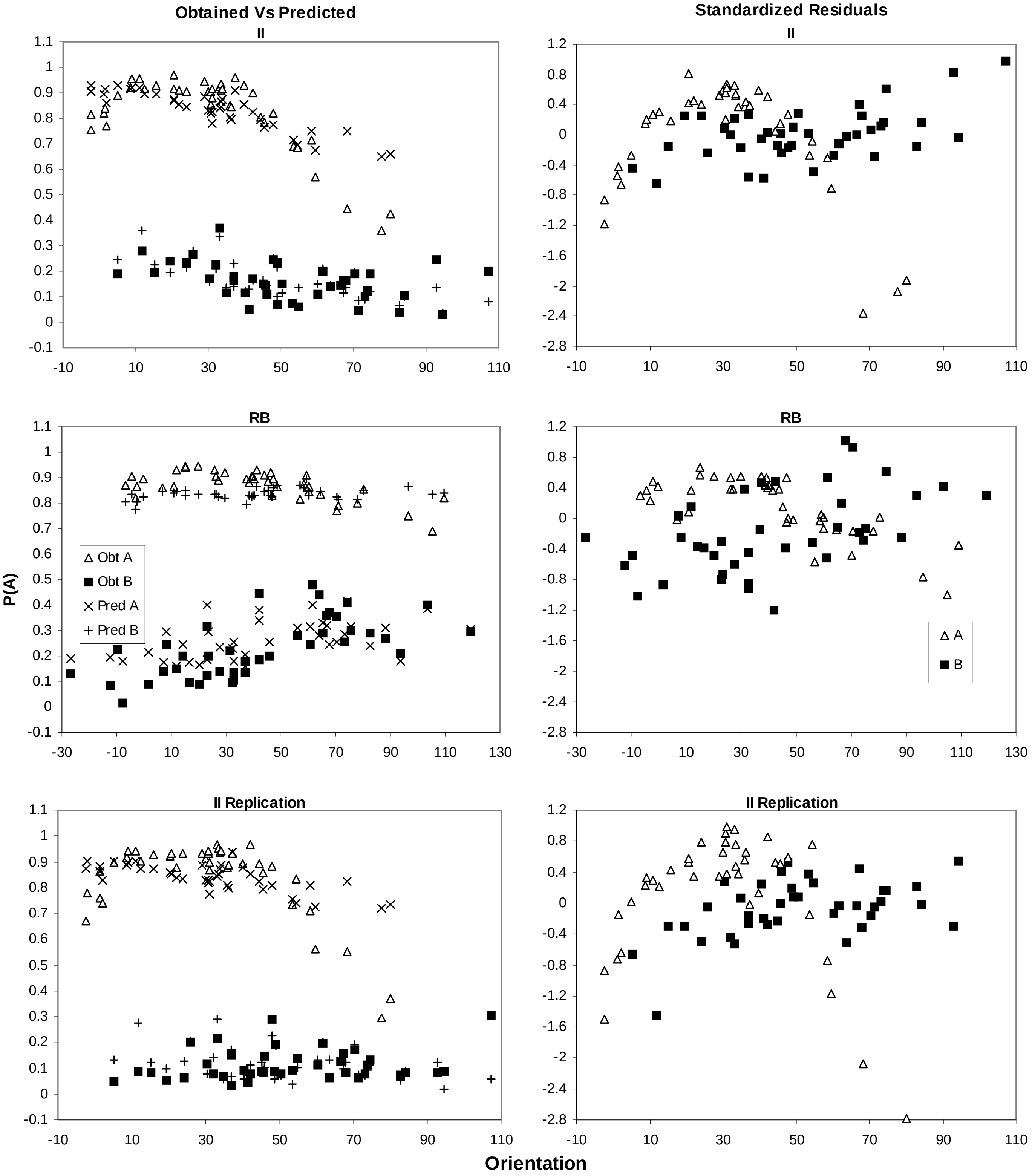


Figure 2.7. Obtained probability of responding category a (P(A) plotted against orientation. The left column shows the obtained data in the unfilled triangles and filled squares for category a and b respectively. Also plotted are the values predicted by the GLC with crosses and x symbols. The right hand column shows the standardized residuals from fitting the GLC.

For the GLC to predict this decrease in accuracy for Category A, the slope of the decision bound must increase, so that the upper part of the line in Figure 2.2 tilts toward the Category A stimuli. But this change in slope means that the decision bound will tilt towards the Category B stimuli for low levels of orientation. This will produce weaker predictions for Category B for low orientation relative to high orientation, which will result in an increase in the residuals for Category B as orientation increases. Thus the significant positive linear coefficients for Category B residuals can be understood, at least in part, as a side effect of the GLC trying to capture the decreasing limb of the inverted-U pattern.

For the RB condition, the GLC is unable to describe the opposing trends evident in the obtained data: *p*(A) decreases with increases in orientation for Category A, but increases with orientation for Category B. This result suggests a perceptual interaction, in which sensitivity of choice responding to differences in frequency in greater at relatively low levels of orientation (near horizontal) than at relatively high levels of orientation (near vertical).

The reason that the GLC is unable to describe the patterns observed in both the II and RB conditions is because it predicts that any linear trend in Category A and B predictions must be correlated. This is because such a trend can only be produced by varying the slope of the decision bound. For example, if the slope in Figure 2.2 decreases, such that predicted *p*(A) for Category A increases as a function of orientation (i.e., strength of prediction for Category A increases), then predicted *p*(A) for Category B must also increase (i.e., strength of prediction for Category B decreases). By contrast, if the slope of the decision bound decreases, then predicted *p*(A) for Category

A will decrease with orientation, and predicted *p*(A) for Category B must also decrease. Thus the fundamental failure of the GLC applied to the present data is that it is unable to predict trends in *p*(A) for Categories A and B as a function of orientation that are not correlated.

DISCUSSION

In Experiment 1 we examined pigeons' performance in a two-dimensional category learning task in which stimuli were dimensionally-separable Gabor patches that varied in terms of their frequency and orientation, similar to those that have been used in research on human category learning (Maddox et al., 2003). We studied two conditions, which differed in terms of whether accurate performance required control by both dimensions ('information integration'; II) or a single dimension ('rule based'; RB). Results showed that pigeons learned both tasks, with an average percentage of correct responses of 85.5% and 82% in the II and RB conditions, respectively. Although perfect performance was possible, responding in all conditions fell short of optimality. Model comparison analyses showed that the General Linear Classifier (GLC; Ashby, 1992), which has been proposed to account for category learning research with humans, provided a better account of responding in the II conditions, but a unidimensional model that assumed control only by frequency provided a better account of results from the RB condition. Results from the II condition confirm that pigeons can pass an empirical test for information integration based on dimensionally-separable stimuli.

Overall we found no difference in accuracy between the II and RB conditions. This was surprising because the RB task in some respects seems intuitively easier – after all, only one dimension is relevant. This finding of near-equivalent performance in the RB and II conditions contrasts with results of similar experiments with humans. For example Maddox and colleagues have found average performance to be 68% for II performance, versus 91% for RB performance (Maddox, Ashby, & Bohil, 2003). However, Ashby, Maddox and colleagues have proposed that humans use two different category learning systems for II and RB tasks. The difference in accuracy for humans under II and RB conditions, relative to the lack of difference with pigeons obtained here, may suggest that pigeons only have a single category learning system.

An unexpected finding was that although the GLC provided a good account of the data overall, with averages of 88% and 85% variance accounted for in the II and RB conditions, respectively, the data deviated systematically from the GLC's predictions. Specifically, we found that the probability of a Category A response (*p*(A)) was an inverted-U shaped function of orientation for Category A stimuli in the II tasks (see Figures 4 and 6). Polynomial regressions on residuals confirmed that these deviations from GLC predictions were systematic (see Table 2.6 and Figure 2.7). This suggests that the GLC is an inadequate model for pigeons' category learning in the II condition. An important goal for future research will be to determine whether similar deviations are obtained with humans responding on II category learning tasks. We report an experiment which tests whether this result is obtained with humans in Chapter 2, as well as a replication with pigeons. If this result is reliable, then a new model for information integration category learning may be warranted.

Systematic deviations from the GLC predictions were also obtained for the RB condition. Performance was better overall at low levels of orientation (close to horizontal) than at high levels of orientation (close to vertical). This result is what would be expected if there were a perceptual interaction between orientation and frequency, such that discriminability of differences in frequency was easier for lines that were near horizontal than for lines that were near vertical. Such an interaction might pose a challenge to the assumption that frequency and orientation are fully separable and independent dimensions. The interaction also might help to explain why pigeons’ performance fell short of optimality in the RB condition. Thus in Experiment 2 we attempted to replicate this result with a different set of pigeons, as well as comparing performance in a rule-based condition in which orientation was the relevant dimension.

EXPERIMENT 2

In Experiment 2 we compared performance in two rule-based conditions, in which either frequency or orientation served as the relevant dimension. There were two major goals: First to assess the adequacy of the GLC and determine whether the systematic deviations obtained in Experiment 1 would be replicated, and whether similar results would be found with orientation as the relevant dimension. More broadly, we were interested in comparing performance in the two RB tasks, and to assess pigeons' performance in terms of optimality. Similar to Experiment 1, we conducted model comparison analyses for each condition to test whether the GLC or unidimensional models provided a better account of the data.

METHOD

Subjects

Three pigeons, designated H5, H6, and H8, participated as subjects and were maintained at 85% of free-feeding weight ± 15 g by post-session feedings. They were housed individually and allowed free access to water and grit, in a vivarium with a 12:12 hr light/dark cycle (lights on at 7:00 a.m.). H5 and H8 had previously served in Experiment 1 and subsequently in an unreported experiment (with H6) using similar apparatus and stimuli, prior to Experiment 2.

Apparatus

Apparatus were identical to Experiment 1.

Stimuli

Two sets of Rule Based (RB) stimuli were produced (Maddox, Ashby & Bohil, 2003). Each of the two stimuli sets can be represented in a 2 dimensional space with orientation on the *x* axis and frequency on the *y* axis. For the RB Frequency condition, the optimal decision bound was a horizontal line drawn through the scatter plot (shown in Figure 3.1), representing a criterial value, such that stimuli with frequencies less than the criterion were assigned to one category, while stimuli with frequencies greater than the criterion were assigned to the other category.

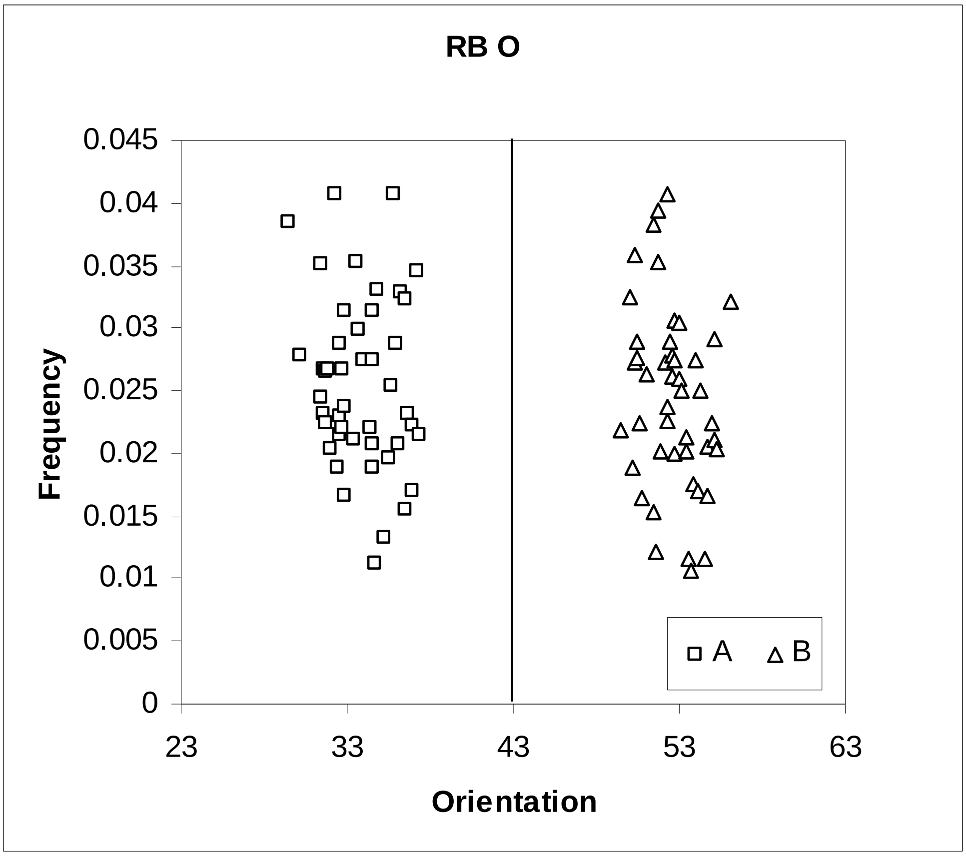


Figure 3.1. Plot based on orientation and frequency on the x and y axis. The optimal decision line is the solid vertical middle line and indicates that orientation is the relevant dimension. Unfilled squares show the distribution of category A and unfilled triangles show category B.

The stimuli for the RB Orientation condition were obtained in the same way as the previous condition except that the criterion decision bound was a vertical line (shown in Figure 3.2) thus making orientation the relevant dimension for successful categorization.

Means and standard deviations, as well as maximum and minimum values for the stimuli in each category for both the RB and II condition are shown in Table 3.1.

	RB F				RB O			
	Cat A		Cat B		Cat A		Cat B	
	Orient	Freq	Orient	Freq	Orient	Freq	Orient	Freq
Min	2.33	0.0270	0.96	0.0207	29.42	0.0112	49.54	0.0107
Max	90.00	0.0297	89.55	0.0229	37.37	0.0408	56.03	0.0406
Diff	87.67	0.0027	88.59	0.0022	7.94	0.0296	6.49	0.0299
Mean	44.91	0.0282	41.61	0.0218	33.89	0.0256	52.63	0.0244
SD	20.51	0.0007	21.88	0.0006	2.04	0.0069	1.64	0.0074

Table 3.1. The characteristics of the stimuli distributions for each category in each condition. The frequency and orientation relevant tasks are indicated by RB F, and RB O, respectively. The left column indicates the characteristic, minimum and maximum orientation and frequency values are in degree and cycles per pixel units. Also shown are the differences between the maximum and minimum values as well as the means and standard deviations.

To generate the RB frequency and orientation condition stimuli, random numbers were sampled from a bivariate normal distribution for each of the categories, A and B. Forty five number pairs were sampled for each category, defining 45 stimuli in terms of frequency and orientation. For display on the LCD screens (640 x 480 resolution), the values were converted to cycles per pixel with the following equation where x equals the randomly generated frequency values: (X/50+.25)/250. The result was a frequency value in cycles per pixel that was maximized for display on a 640 x 480 screen. For the orientation dimension, numbers were converted to degrees from horizontal by first multiplying by π/500, then to degrees from horizontal by multiplying the radian value by 180 then dividing by π.

Gabor stimuli were generated in real-time using custom software written at the University of Canterbury. The algorithm used was based on the Gabor Filter, (Yao, Krolak, & Steele, 1995), and was integrated into a C++ program that displayed the images based on a pre-determined CSV file listing of frequency (cycles per pixel) and orientation (degrees).

Procedure

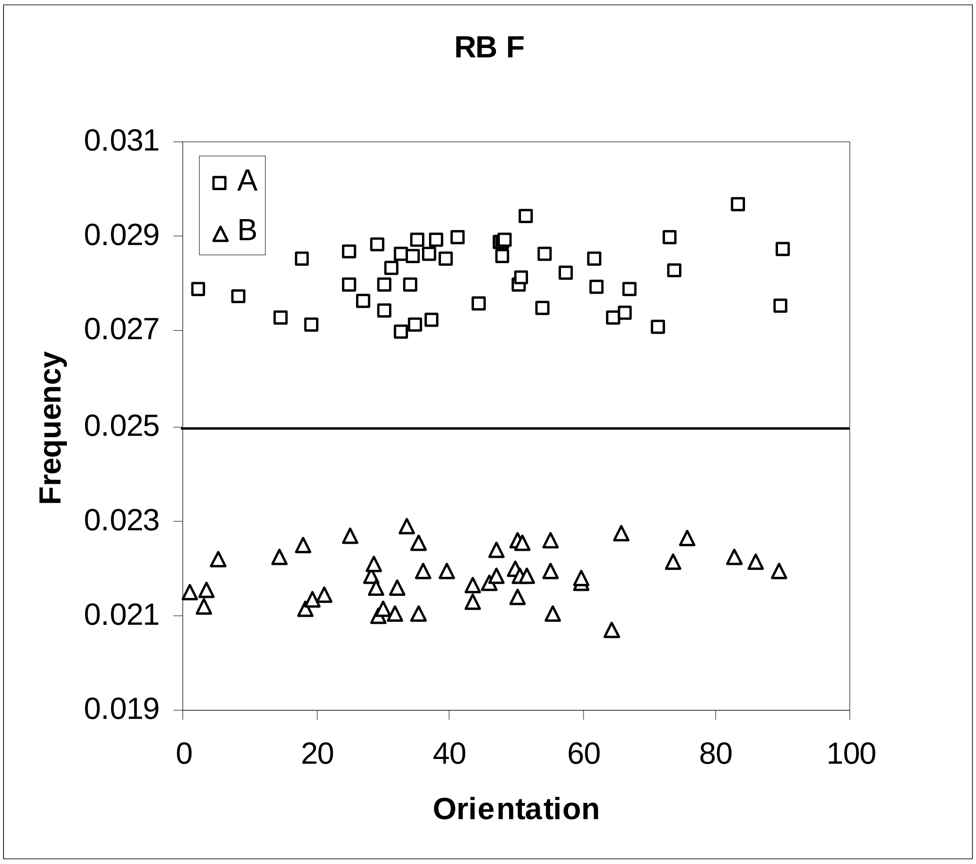


Figure 3.2. The optimal decision line is the solid horizontal middle line and indicates that frequency is the relevant dimension.

All pigeons had previous experience pecking the touch screen and side response keys in other categorization procedures, and so training began immediately on the final procedure, which was the same as that used in Experiment 1.

	Condition Order	
H5	Frequency	Orientation
H6	Frequency	Orientation
H8	Orientation	Frequency

Table 3.2. The order of conditions for each bird are indicated by the relevant dimension.

Pigeons were exposed to the RB frequency and orientation conditions in counterbalanced order. Training continued in each condition until all birds had reached a visual stability criterion. The keys assigned to the categories, correct key location and color were counterbalanced across birds and are listed in Table 3.2, along with the order of conditions used in training.

RESULTS

Figure 3.3 displays the percentage of correct responses across all sessions in the experiment and Table 3.3 contains the mean and standard deviations for each bird in each condition. All subjects achieved high degrees of accuracy in both conditions. Subjects H5 and H6 had the highest overall performance in both conditions, with accuracy levels that approached optimality for the RB-frequency condition (0.96 and 0.95, respectively). Mean and standard deviation for percentage correct responses are listed for all subjects and conditions in Table 3.3. Overall, accuracy levels were not significantly different in the frequency and orientation conditions, *M*'s = 0.92 and 0.89, respectively, *t*(2) = 1.02, *ns*.

	Frequency		Orientation	
	Mean	SD	Mean	SD
H5	0.96	0.020	0.89	0.037
H6	0.95	0.020	0.92	0.091
H8	0.84	0.091	0.86	0.034
Overall	0.92	0.044	0.89	0.054

Table 3.3. Mean and standard deviations for each bird's performance in each condition as well as the overall average displayed in the bottom line.

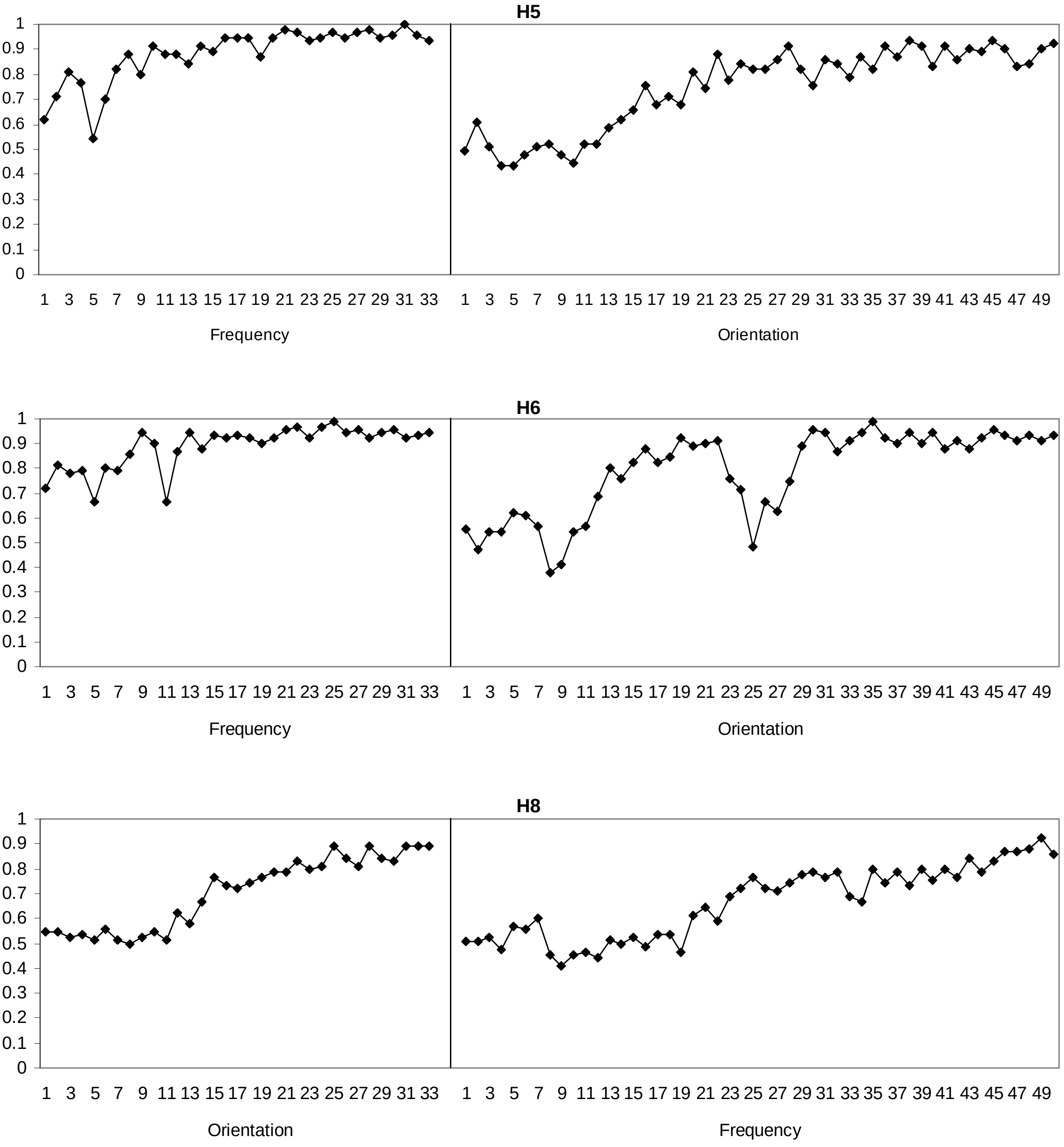


Figure 3.3. Accuracy plots for each bird and each condition. Frequency relevant and orientation relevant performance are plotted from left to right. Session number is displayed on the x axis and percent correct is on the y axis.

Model Comparison

The GLC and two unidimensional models (UNI-F and UNI-O) were fitted to the individual data from both conditions using the same maximum-likelihood estimation procedure as in Experiment 1, and compared using the AIC. Table 3.4 shows the AIC and variance accounted for (VAC) for each subject and condition. The best-fitting model is indicated by the lowest AIC for each data set, displayed in boldface.

	GLC		UniF		UniO	
	VAC	AIC	VAC	AIC	VAC	AIC
H5 Freq	0.98	38.06	0.98	**34.10**	0.01	127.96
H5 Orient	0.93	66.25	0.00	128.42	0.93	**62.41**
H6 Freq	0.96	41.33	0.96	**37.36**	0.01	127.76
H6 Orient	0.95	59.25	0.01	127.78	0.95	**55.38**
H8 Freq	0.87	81.28	0.87	**77.32**	0.01	126.13
H8 Orient	0.91	77.24	0.05	125.87	0.89	**74.81**

Table 3.4. Akaike Information Criterion (AIC) values as well as variance accounted for (VAC) results for each subject and each condition.

In the frequency condition, the UNI-F model had the lowest AIC value for all subjects. Conversely, the UNI-O model had the lowest AIC value for all subjects in the orientation condition. This shows that responding was controlled by the single relevant dimension in both conditions. Note that the GLC also provided a good fit in terms of variance accounted for; this is not surprising because the UNI-F and UNI-O models represent a special case of the GLC, which can mimic the unidimensional decision bounds associated with both models.

	GLC			UNI-F		UNI-O	
	Slope	Intercept	Noise	Mean	Sigma	Mean	Sigma
H5 Freq	0.0000	0.0249	0.0017	0.0247	0.0017	37.6629	182.4940
H5 Orient	-0.0144	0.6569	0.1034	0.0206	0.1078	43.9697	7.2336
H6 Freq	0.0000	0.0242	0.0017	0.0244	0.0017	26.9371	200.3907
H6 Orient	0.0157	-0.6495	0.1046	0.0259	0.0541	43.0476	6.6755
H8 Freq	0.0000	0.0262	0.0029	0.0260	0.0029	83.0354	207.4903
H8 Orient	0.0040	-0.1526	0.0333	0.0220	0.0336	44.5999	8.4266

Table 3.5. Model parameters for the general linear classifier (GLC) and each of the unidimensional models (Uni-F and Uni-O) for each bird in each condition.

Table 3.5 shows the estimated parameters for each of the three models. Parameter values for the best-fitting models in each condition were consistent across subjects and indicated that subjects set their criterion values at optimal or near optimal values. For example in the UNI-F condition, the average estimated mean was .0250, which was equal to the average of the frequency values for all stimuli across both categories, indicating that subjects placed their criterion at the midpoint of the stimulus range. Similarly, the average estimated mean for UNI-O model in the orientation condition was 43.87, which was close to the average orientation value across categories (43.26).

Detailed Analyses of Asymptotic Performance

Next we examined asymptotic performance in greater detail by plotting the probability of responding Category A (P(A)) for all stimuli as a function of the category-irrelevant dimension in each condition. Plots were made against orientation in the frequency condition, and vice versa, to examine whether there was any control over responding by the irrelevant dimension.

Results for the frequency condition are shown in Figure 3.4. The left column plots P(A) as a function of orientation. Although there was variability across subjects, there was some evidence of an interaction for Pigeons H6 and H8, similar to that obtained in Experiment 1. For these subjects, P(A) decreased for Category A stimuli as orientation increased, whereas P(A) increased for Category B stimuli. This suggests that accuracy decreased overall for high levels of orientation (i.e., near-vertical grating). By contrast, results for H5 showed no such

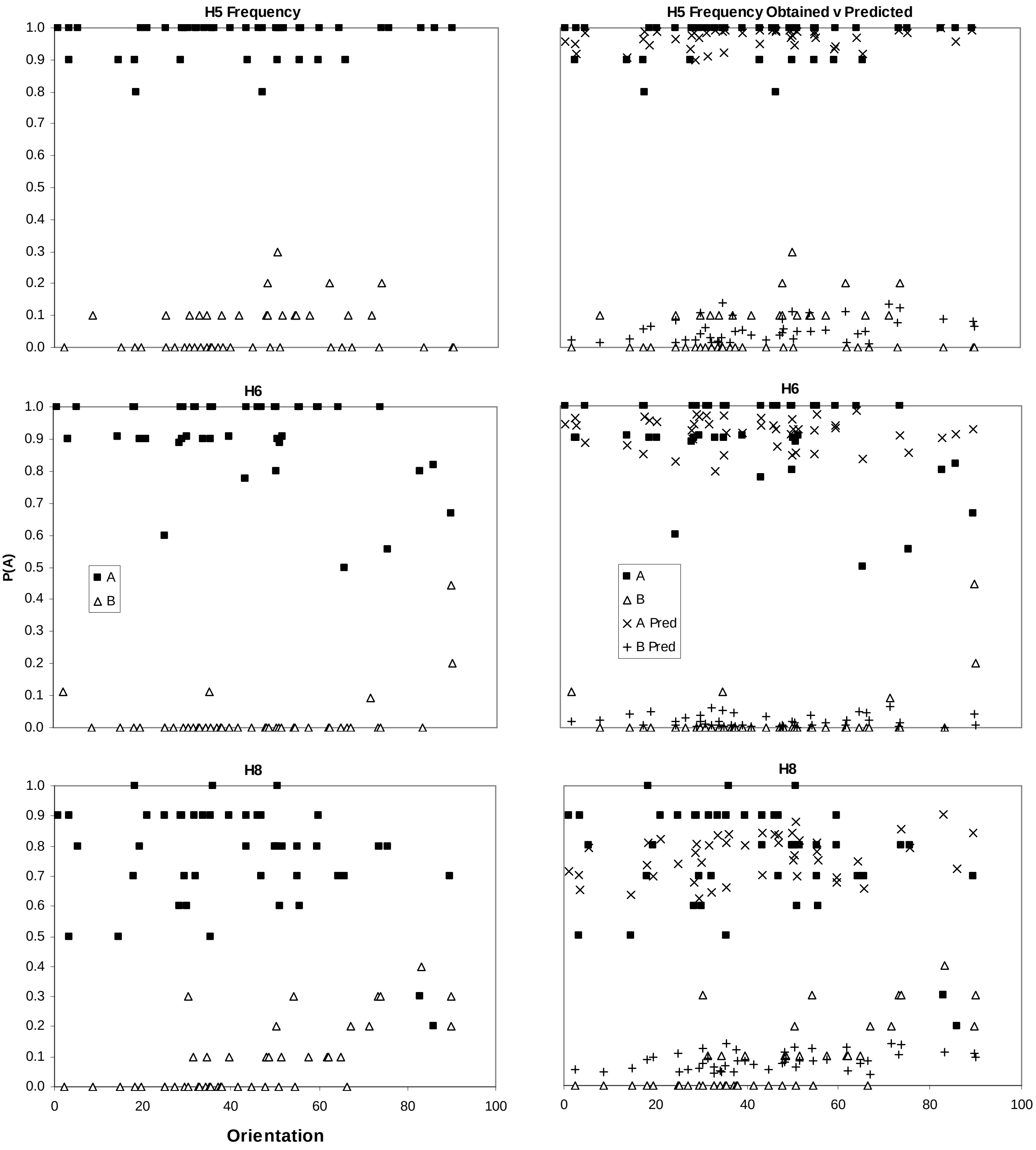


Figure 3.4. Probability of a category A response (P(A) in the y axis plotted against orientation in the x axis. The right column contains obtained and predicted values for comparison. Category is A represented by the filled squares and B the unfilled triangles.

pattern – accuracy levels were consistently high across the range of orientation values for both categories. The right hand column of Figure 3.4 shows the same obtained P(A) plots as the left column but also have the predicted values of the GLC shown as well. Note that the GLC predictions always approximate two horizontal lines. The lines are parallel, suggesting that the GLC cannot predict the interaction in the data for H6 and H8.

Figure 3.5 shows P(A) for the orientation condition plotted as a function of frequency values. Contrary to results from the frequency condition, there was no clear evidence of an interaction in the data: For all subjects, there was no apparent trend in accuracy as a function of orientation. The obtained P(A) and predicted plots shown in the right column once again show that the GLC predicted parallel linear functions for each category.

Average data for the frequency condition are displayed in Figure 3.6. The upper panel shows P(A) as a function of orientation, and predictions of the GLC are added in the middle panel. The lower panel shows the standardized residuals from the GLC fit. The interaction evident in the results for H5 and H8 is apparent in the average data: Accuracy of responding is lower for stimuli with high orientation values, and this effect cannot be accounted for by the GLC. Figure 7 shows a similar summary of results from the orientation condition. By contrast, there was no evidence of an interaction, and no systematic pattern in the GLC residuals.

Regression Analyses

As in Experiment 1, we conducted a series of regression analyses on the residuals from the GLC fits to determine whether the deviations from the model's fits were systematic. Table 3.6 displays the results of these analyses for both RB-orientation and RB-frequency conditions. The three columns on the left show results for Category A; the three columns on the right show results for Category B. Beta weights (β) for both linear and quadratic components and R^2 values are displayed. Table 6 shows that for the RB-frequency condition, linear coefficients were always negative for Category A and positive for Category B, and reached significance for two individual pigeons (Category B), as well as for the group mean data. This pattern is consistent with results from the RB-frequency condition in Experiment 1. There was also a significant positive quadratic component for H6 and H8 for Category B (as well as for the mean data), and a significant negative quadratic component for the mean data for Category A. The negative quadratic for Category A is also similar to Experiment 1, and suggests that accuracy for Category A (the upper one in the figures) tended to show an inverted-U shaped pattern.

By contrast, few coefficients were significant for the RB-orientation condition. The only significant result was a positive quadratic coefficient for H6 for Category B, which also was obtained in the mean data. This component suggests there was some tendency for accuracy for Category B (the lower one in the figures) to show a U-shaped pattern. Although this result was not obtained consistently across pigeons, it was qualitatively similar to results from the RB-frequency condition in that it suggested that performance declined at the extreme levels of the irrelevant variable (in this case, frequency).

	Category A			Category B		
Bird, Cond	β Linear	β Quadratic	R^2	β Linear	β Quadratic	R^2
H5, Freq	-0.12	0.08	0.01	0.04	-0.29	0.08
H6, Freq	-0.26	-0.34	0.21	0.25 *	0.60 ***	0.48 **
H8, Freq	-0.41	-0.37	0.34 **	0.26 *	0.60 ***	0.48 **
Mean, Freq	-0.42 ***	-0.40 **	0.38 **	0.25 *	0.48 **	0.34 *
H5, Orient	-0.09	0.22	0.05	-0.21	0.10	0.05
H6, Orient	0.24	-0.24	0.09	-0.22	0.55 ***	0.31
H8, Orient	-0.14	0.08	0.02	-0.18	0.11	0.04
Mean, Orient	-0.01	0.03	0.00	-0.27	0.36 *	0.18

Table 3.6. Regression analysis results for category A and B stimuli. Mean values are shown on the bottom line. Each category reports the Beta Weights (β) for the linear coefficient, β for the quadratic coefficients and the R^2 values. Significant regression results for β Linear or β Quadratic are indicated with an asterisk (*) for 0.1, 0.05, 0.01 significance levels using *, ** and *** respectively.

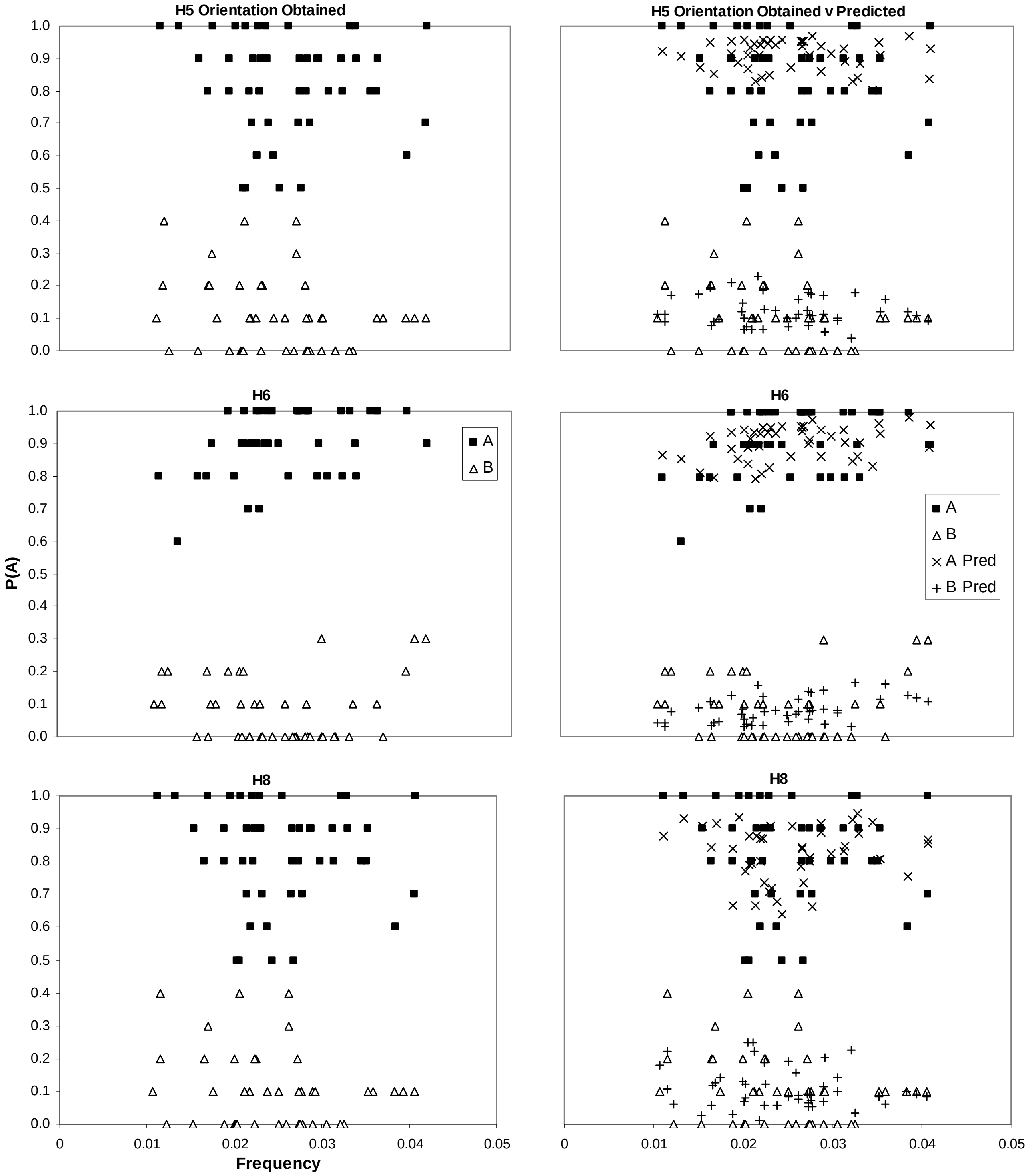


Figure 3.5. Probability of a category A response (P(A) in the y axis plotted against frequency in the x axis. The right column contains obtained and predicted values for comparison. Category is A represented by the filled squares and B the unfilled triangles.

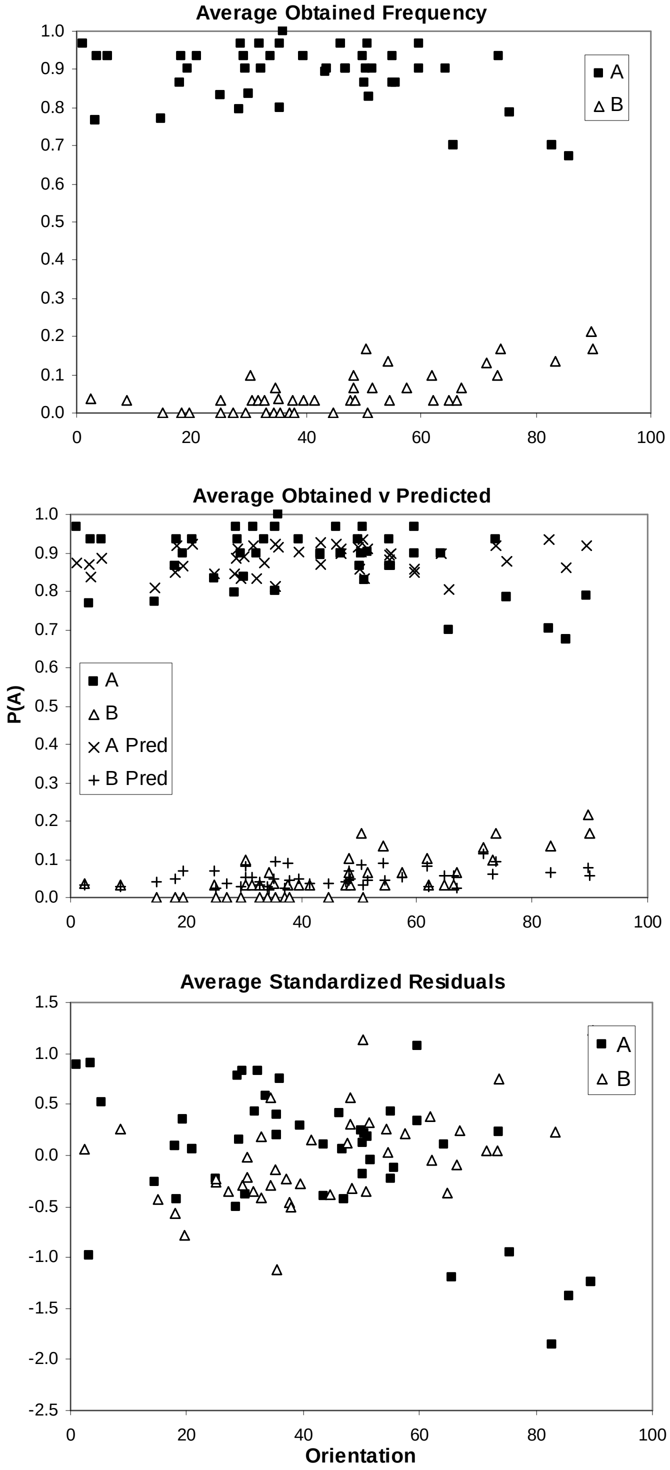


Figure 3.6. Overall average obtained values are contained in the top panel. The middle panel shows the average obtained vs. predicted values and the bottom panel displays the average standardized residuals from the GLC model fits.

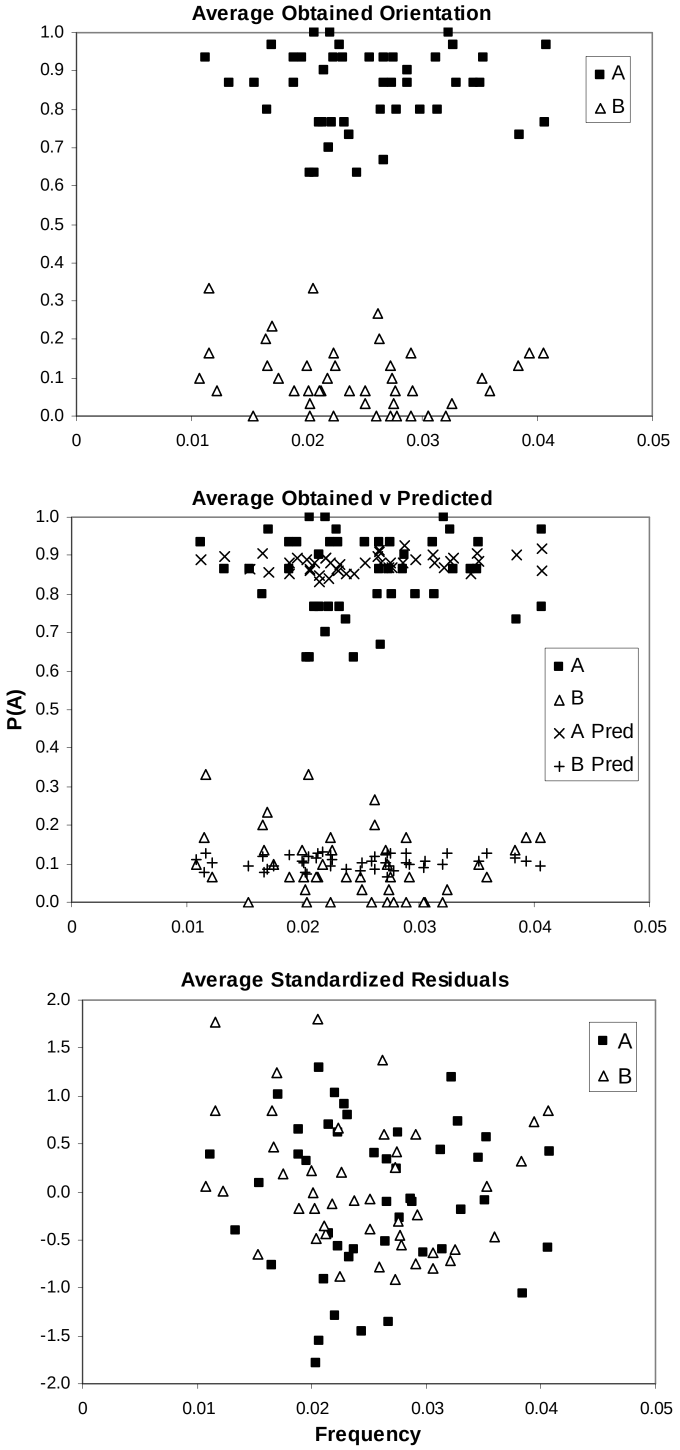


Figure 3.7. Overall average obtained values are contained in the top panel. The middle panel shows the average obtained vs. predicted values and the bottom panel displays the average standardized residuals from the GLC model fits.

DISCUSSION

Experiment 2 compared performance in two rule-based conditions in which either orientation or frequency was the relevant dimension. Results showed that all pigeons learned both discriminations to a high degree of accuracy, although performance did not reach 100% accuracy. Of particular interest was whether the interaction obtained in Experiment 1, in which performance in the RB-frequency task was more accurate at orientation levels close to horizontal rather than vertical, would be replicated, and whether it would also be obtained when orientation was the relevant dimension. We found that for two pigeons, residuals from fits of the GLC showed systematic deviations similar to those obtained in Experiment 1. For the RB-orientation task, the evidence for an interaction was not as strong; the linear trend in the residuals was not significant, although there was some indication for at least one pigeon (H6), that accuracy for Category B was less at extreme levels of frequency.

Interestingly, the pigeon that showed no interaction in the RB-frequency condition, H5, also had the highest performance, achieving near-optimal levels of accuracy (0.96). This makes sense because by definition, optimal responding would imply that performance was constant regardless of orientation. Thus we obtained a similar interaction in the RB-frequency condition as Experiment 1, when accuracy levels were less than optimal. Results for the RB-orientation condition were less clear, although for one pigeon performance was significantly better for mid-range frequency values. The significant quadratic components obtained in the RB-frequency condition here and in Experiment 1 were also consistent with the general view that performance was less affected by the irrelevant dimension when values on that dimension were towards the middle of the range.

Accuracy levels were also overall higher than the comparable condition of Experiment 1. Pigeons H5 and H8 had previously served in Experiment 1, where their accuracy for the RB-frequency condition was 0.84 and 0.78, which compares with 0.96 and 0.84, respectively, in Experiment 2. Although responding appeared to be stable in both experiments, it is possible that improved performance occurs when pigeons are given long-term exposure to the procedure.

GENERAL DISCUSSION

The primary goal of this study was to examine performance of pigeons in a two-dimensional visual category learning task derived from Ashby and Gott's (1988) randomization procedure. Previous research by Herbranson et al. (1999) showed that pigeons could respond at near-optimal levels in category tasks that required attention to two dimensions ('information integration') or attention to a single dimension. However, unlike Herbranson et al.'s (1999) study, which used rectangles that varied in height and width as stimuli, we used Gabor stimuli. The advantage of Gabor stimuli is that their component dimensions, frequency and orientation, are fully separable. One potential problem with Herbranson et al.'s study was that because height and width are measured in the same units and thus not fully separable, pigeons might have been able to respond accurately in the information integration condition by comparing whether the height was greater than the length, or vice versa. This difficulty is avoided with Gabor stimuli, and thus the present study represented a more conservative test of pigeons' capacity for information integration. Another difference between our study and Herbranson et al. is that we used non-overlapping category distributions, such that perfect performance was possible. Because this raises the level of performance that subjects could attain, it arguably provides a stronger test of optimality.

In Experiment 1 we examined pigeons' responding in both an information integration condition and a selective attention or 'rule based' condition for which frequency was the only relevant dimension. Results showed that pigeons could respond accurately in both conditions, although performance fell short of optimality. Model comparison analyses confirmed that the General Linear Classifier (Ashby, 1992), a decision bound model for information integration, provided a better account of performance than unidimensional models that assumed responding was controlled by either frequency or orientation. This confirms that pigeons can pass an empirical test for information integration with fully-separable stimuli: Their category decisions were accurate and depended on both frequency and orientation.

Our failure to find optimal performance contrasts with results of Herbranson, Fremouw and Shimp (1999). Herbranson et al. found that their pigeons performed nearly optimally when categorizing rectangular stimuli that varied in terms of height and width. In their procedure, stimulus categories were overlapping bivariate normal distributions and perfect performance was impossible. Nevertheless, Herbranson et al. found that performance was close to that predicted by an optimal linear decision bound.

There were several procedural differences between Herbranson et al.'s study and the present experiment that might account for the different results. Two have already been mentioned – the use of rectangular stimuli and overlapping category distributions in Herbranson et al.'s study, compared to Gabor stimuli and non-overlapping category distributions used here. However, there is no apparent reason why either of these factors would affect the overall degree of performance. Another possibility is that our task may have been more difficult than Herbranson et al.'s because stimuli from the two categories were closer together, relative to the variability within- and between categories. To investigate this possibility, we calculated effect sizes for the distance between category centroids for the II condition in both Herbranson et al. and our Experiment 1. Specifically, effect size was defined as the Euclidean distance between the centroids of Category A and B, divided by the pooled standard deviation. For Herbranson et al., the effect size was 3.29, whereas for the present study the effect size was 1.33. This means that the categories in our study, although not overlapping, were on average closer together than those in Herbranson et al., relative to the stimulus variability. The implication is that our category tasks may have been more difficult than Herbranson et al., which would account for the suboptimal performance.

However, closer inspection showed that the data deviated systematically from the predictions of the GLC. Specifically, accuracy for Category A stimuli (i.e., the upper category in Figure 2.2)) was an inverted-U shaped function of orientation (see Figure 7). Polynomial regressions confirmed that this pattern was significant. By contrast, accuracy for Category B stimuli (i.e., the lower category in Figure 2.2) did not vary systematically with orientation. These results were also obtained when the information integration condition was replicated.

The inverted-U shaped pattern for Category A stimuli may be related to the pigeons’ suboptimal performance, because it was associated with decreased accuracy for orientations that were outside the middle range and were close to horizontal or vertical. Exactly why this pattern was obtained is unclear. However, the fact that it was found with six pigeons and replicated suggests that it is reliable. It may represent a kind of ‘footprint’ for how pigeons learn the information-integration task. An important question is whether human performance under similar conditions would show the same result. If so, then a new model for category learning may be warranted. Testing this possibility was the major goal of the research reported in Chapter 3.

Chapter 3 – Comparing information-integration category learning by pigeons and humans

Results from Chapter 2 suggested that performances of pigeons in the information-integration task deviated from optimality and the predictions of the GLC in a systematic way. The experiments reported in Chapter 3 had two primary goals. First, we wanted to see if the systematic deviations (i.e., inverted-U shaped accuracy pattern for Category A responding) would be obtained with a different set of pigeons and arrangement of stimulus values. The experiments reported in Chapter 2 used bivariate distributions from Maddox et al. (2003) to generate stimuli. The orientation values were approximately distributed around 45 degrees (the averages for Category A and B were 52 and 31 degrees, respectively), but there were a few exemplars which had negative orientation values. To ensure that results were not affected by having some stimuli with near-zero (i.e., horizontal) or negative (i.e., lines going down from left to right) orientations, in the experiments reported in Chapter 3 the stimuli were centered on 45 degrees and varied within an 80 degree range (i.e., from 5 degrees to 85 degrees). Our second major goal was to test whether a similar pattern of deviations from GLC predictions would be obtained with humans. Thus Chapter 3 includes two experiments which directly compare pigeons' (Experiment 1) and humans' (Experiment 2) performance in the information-integration task. A secondary aim was to test for stimulus range effects in the information-integration task, specifically whether limiting the range of orientation values would affect control by that dimension.

INTRODUCTION

Categorization is an essential ability for the survival of any organism. Animals as well as people make many category judgments everyday, in many cases with great accuracy. Consider a dogsled racing team, which is typically composed of 16 dogs and 1 human musher. The team must be continually watching and evaluating the terrain and the climate while traveling. Dogs communicate with the musher using body language and different barks or yelps. This enables them to warn the musher of hazards on the trail ahead, changes in terrain or problems with their health. The musher in turn must learn to interpret the signals the dogs give and also evaluate the conditions for themselves. All of this information that is processed by both the dogs and also the musher must be categorized correctly in order for the team to function effectively in a dogsled race (Sherwonit & Schultz 1991). Because categorization is such a fundamental skill in all organisms, understanding how this is accomplished is an important task. Given the evidence that many phenomena in conditioning and learning have generality across species, the question arises as to whether a common set of underlying mechanisms can account for categorization performance in both humans and nonhumans. The current research sets out to investigate that question through comparative study, and aims to identify possible common processes that might have evolved for category learning.

Studies that have compared human and nonhuman performance on identical category learning tasks have been relatively rare. A notable recent study was conducted by Smith, Minda, and Washburn (2004), who compared performances of humans and rhesus monkeys (*Malacca mulatta*) in the categorization tasks that were used in classic research by Shepard, Hovland and Jenkins (1961).

In Shepard et al.'s study, humans categorized stimuli that varied on three binary dimensions: shape (triangle vs square), size (large or small), and color (black or white). Thus there was a total of eight different stimuli, which were divided into two categories with four stimuli each. Shepard et al. arranged six different groupings of the stimuli, yielding six categorization tasks. The tasks were: I: criterial attribute (in which color was the single diagnostic dimension); II: exclusive-or (XOR) or correlated features, in which for example black triangles and white squares were in Category A, while black squares and white triangles were in Category B; tasks III and V were 'rule plus exception', in which for example, black stimuli were always in Category A and white stimuli in Category B, except for the small white triangle and small black square; task IV was a family-resemblance or prototype task which was similar to the rule-plus-exception tasks in that stimuli within a category shared values on two of the three dimensions; and task VI was a polymorphous task in which no features were diagnostic. The critical question posed by Shepard et al. was whether performances could be explained in terms of the basic behavioral processes of conditioning and generalization, which predicted that difficulty should vary across the tasks depending on the degree of perceptual similarity of the stimuli between-categories. According to this view, the easiest tasks should be those with high within-category similarity and low between-category similarity. However, Shepard et al. found that results did not support the view that humans learned the category tasks through a gradual conditioning process. Instead, humans performed much better on the Type II (XOR) task, and in general, responding showed abrupt shifts in accuracy. Shepard et al. concluded that humans were solving the problems through testing of explicit rules and hypotheses, rather than a gradual conditioning process.

Smith et al. (2004) attempted to replicate Shepard et al.'s tasks with both humans and rhesus monkeys. They found that performance of monkeys, contrary to the humans, was consistent with basic conditioning processes: Task difficulty varied directly with the perceptual similarity of the stimuli. The most striking difference between humans and monkeys was obtained with the Type II (XOR) task, which monkeys found very difficult, but humans were able to learn very quickly. Smith et al. concluded that their results were consistent with the general view that humans could learn categories via explicit hypothesis testing or through a gradual conditioning process (Ashby et al., 1998), whereas monkeys appeared to be limited to the latter system. When the problem could not be stated in terms of a simple rule (i.e., Task VI), humans learned only gradually. Thus Smith et al.'s results suggested that humans could use an explicit, rule-based system or a non-verbal information integration system.

Cook and Smith (2006) compared performance of pigeons and humans on a 'rule plus exception' task, similar to Smith and Minda (1998). In this task, subjects categorized circles comprised of six wedges, each of which can be one of two colors. The task is structured so that each category contains a prototype (most typical member), three exemplars which are similar to the prototype but differ in terms of one of the colors, and an exception which is more similar to the other category. The critical result reported by Smith and Minda (1998) was that early in training, humans would categorize the exception as belonging to the other category. However, given sufficient training, they would learn to respond correctly to the exception. They concluded that humans used an abstraction process (similar to prototype theory) to learn the task, but later switched to an exemplar-based memorization process. Cook and Smith (2006) found that pigeons showed exactly the same pattern of results.

In Shepard et al.’s (1961) influential study and much subsequent research on categorization, the stimuli were multidimensional, but only a limited number of values were used on each dimension. However, a paradigm developed by Ashby and Gott (1988) allows for a potentially infinite number of stimuli to be used that vary parametrically on two dimensions. Specifically Ashby and Gott (1988) had participants categorize L-shaped line segments that differed in terms of the heights of the vertical and horizontal lines. To generate stimuli, Ashby and Gott used a randomization procedure in which categories were defined in terms of two bivariate normal distributions. The category dimensions were two dimensional (height and width), and the distributions that formed the categories overlapped such that perfect performance was impossible, but an optimal decision bound was formed by a diagonal line running through the stimulus space with a slope of approximately 1. Results showed that performance of the three subjects was close to optimal with an average accuracy of 83%. The authors described their results in terms of Ashby and Townsend’s (1986) General Linear Classifier (GLC), which is effectively a generalization of signal detection theory.

Herbranson, Fremouw and Shimp (1999) examined whether pigeons could also respond optimally in the randomization task introduced by Ashby and Gott (1988).

They trained pigeons to categorize rectangles displayed on a computer screen that varied in terms of height and width and were generated using two bivariate normal distributions that overlapped. In the ‘divided attention’ condition, accurate performance depended on both dimensions: Rectangles for which the height was greater than the width were likely to belong to Category A, whereas rectangles for which the width was greater than the height were likely to belong to Category B. In a second condition, ‘Selective Attention’, accurate performance depended on only one dimension. For example wide rectangles might belong to one category and narrow to the other, but the height of the rectangles was irrelevant. Results showed that the pigeons’ performances in the tasks were close to optimal in both tasks. These results suggest that pigeons are capable of integrating information from two relevant dimensions and respond optimally in a two-dimensional information integration task, similar to humans.

Thus previous research has demonstrated that performances of humans and nonhumans are similar when the categories cannot be learned in terms of verbal rules, and that both humans and pigeons can respond optimally on Ashby and Gott’s (1988) information-integration task. However, our results in Chapter 2 pose some questions. We found that pigeons’ performances in the information-integration task were not only suboptimal, but also deviated from optimality in a systematic way that could not be explained by the GLC. The first question is whether these results were reliable, or might have been due to idiosyncratic features of our research design. We used stimuli that were modeled after Maddox, Ashby and Bohil (2003), in which the range of orientation values was greater than 90 degrees. Thus some stimuli had orientations that were less then 0 degrees (meaning subhorizontal) while some were greater than 90 degrees (meaning subvertical). Because there is only 180 degrees of range in Gabor stimuli (i.e. 0 degrees = 180 degrees), stimuli with orientations less than 0 or greater than 90 degrees may be more perceptually similar than stimuli with orientations equal to 0 and 90 degrees. Thus, we wanted to attempt to replicate the results from Chapter 2, except using stimuli that varied within an 80 degree range, centered on 45 degrees.

Second, we wanted to compare performance of pigeons and humans on the information-integration task. If the pattern of suboptimal results was replicated for pigeons with the 80 degree range, then we wanted to know whether similar results would be obtained with humans. As noted above, previous research has generally found that performances of humans and nonhumans are similar for category learning tasks that cannot be described in terms of verbal rules. If we find that the results from Chapter 2 are replicated here with humans as well, then that could have major implications for our understanding of category learning. For example, it would suggest that the GLC is inadequate as a descriptive model for performance in the information-integration task, and that a new model may be warranted.

Finally, as a secondary aim, we planned to examine the effects of stimulus range in categorization tasks. In human subjects past research on stimulus range has formulated the Range-Frequency Theory (Parducci & Wedell, 1986). This theory specifies a ‘range principle’ which essentially states that as the range of stimulus values on a particular dimension increases, the sensitivity to changes along that dimension decreases, and vice versa. Research into range effects has found similar results, namely that as a stimulus range narrows sensitivity within the range increases and the categorization task becomes easier (Hinson & Lockhead, 1986; Lockhead, 2004). Notably these range effects have been found using categorization tasks that involve single dimensional tasks. Our research sets out to test whether the ‘range principle’ applies in the case of a two-dimensional task as well.

Thus we report two experiments which investigate performance in the information integration task with Gabor stimuli by pigeons (Experiment 1) and humans (Experiment 2). In both experiments, subjects were exposed to two conditions in which the overall range of orientation values was Wide (80 degrees) or Narrow (10 degrees).

EXPERIMENT 1

METHOD

Subjects

Three pigeons, designated G6, G7 and G8, participated as subjects and were maintained at 85% of free-feeding weight ± 15 g by post-session feedings. They were housed individually and allowed free access to water and grit, in a vivarium with a 12:12 hr light/dark cycle (lights on at 7:00 a.m.). All were experimentally naïve with no previous experience.

Apparatus

Three operant chambers, 350 mm deep by 360 mm wide by 350 mm high, were used. One wall contained an aluminum response panel in which a VGA 6.4” (130 mm wide x 97 mm tall) LCD display set to 640 x 480 resolution was mounted The LCD display was located 165 mm from the side edge and 230 mm from the bottom floor to center of the screen. Overlaying the LCD screen was a glass panel mounted resistive touch screen of identical size to the screen with a 4096 x 4096 point array resolution. Screen responses were measured via a USB touch interface (Elo TouchSystems Inc). The displays with touch panels were purchased from Touch Screens Inc, part number MTF064D. There were two vertically-aligned response keys on each side of the screen, midway between the edge of the screen and the chamber wall. The keys were 25 mm in diameter, and could be illuminated with 5 color LED arrays. A force of approximately 0.10 N was necessary to operate each key, and produced an audible feedback click. Centered below the screen was a grain magazine with an aperture (60 mm by 50 mm) 40 mm above the floor. The magazine was illuminated when wheat was made available by a white LED. A houselight was centered above the LCD screen 10 mm from the top of the panel. Chambers were enclosed in a sound-attenuating box, and ventilation and white noise were provided by an attached fan. Event scheduling, data recording, and screen image display was controlled with an IBM®-compatible microcomputer. Chamber keys, grain magazine and all other hardware inputs and outputs were interfaced via a USB module with 24 bits of digital I/O purchased from

Measurement Computing (part # USB-1024LS).

Stimuli

The stimuli for the categorization tasks were Gabor patches. Gabor patches are sine wave gratings modulated by a circular Gaussian filter, and vary in terms of frequency and orientation. Sample Gabor patches are shown in Figures 4.1 and 4.2. Two sets of Gabor stimuli were produced to yield two different experimental conditions (Maddox, Ashby, & Bohil, 2003). Each of the two stimuli sets can be represented in a two-dimensional space with orientation on the X axis and frequency on the Y axis. The Wide 80° condition had orientation values with a range of 80° centered around 45° from horizontal (5° to 85°) while the Narrow 10° condition had a range of 10° also centered around 45° (40° to 50°). Pilot testing showed that the differences between the 40° and 50° stimuli in the Narrow condition were easily discriminable to a human observer.

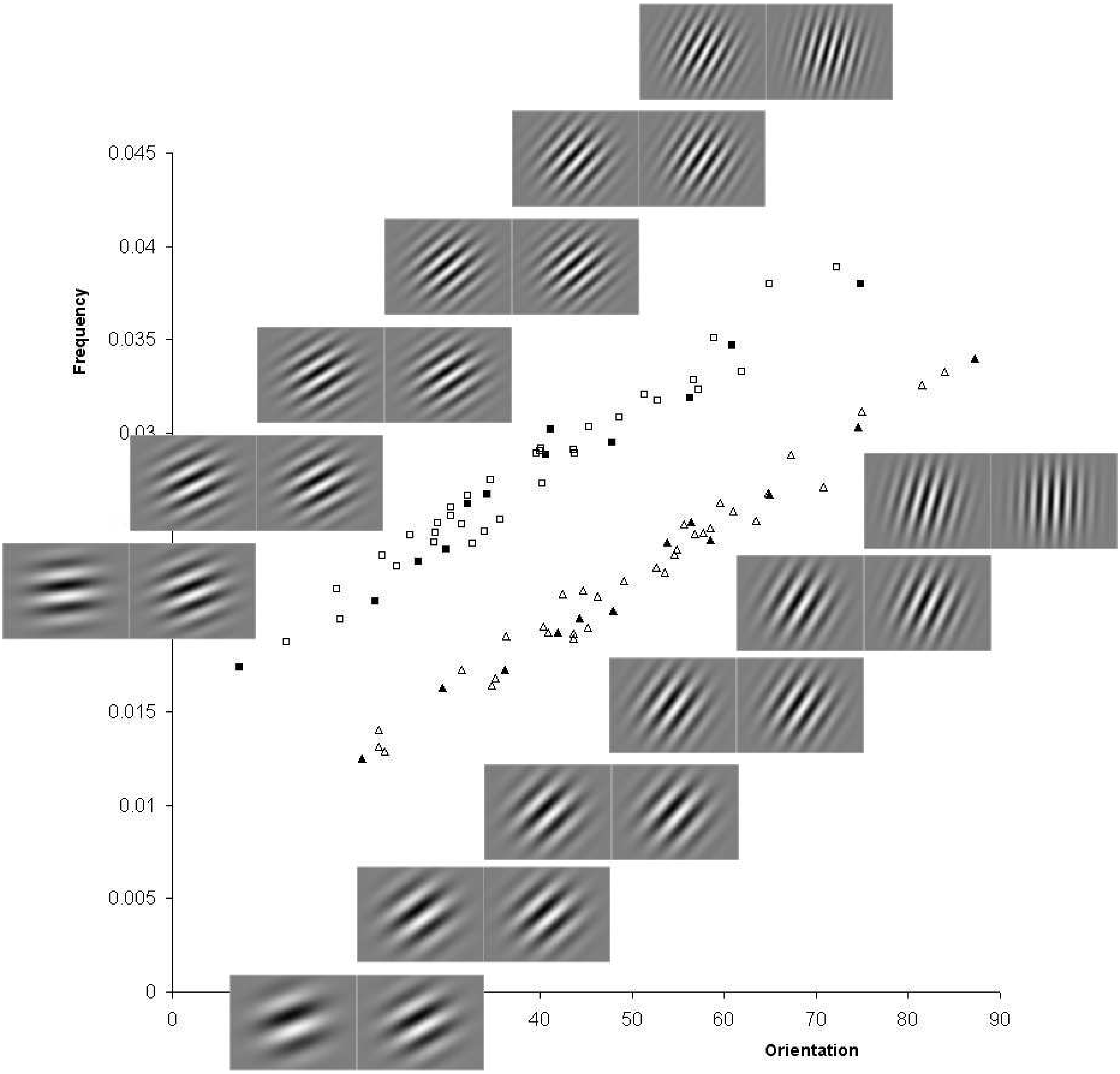


Figure 4.1. The stimuli distribution of the Wide condition, X axis orientation values are in degree units and y axis frequency values are in cycles-per-pixel units. The unfilled squares show category A stimuli and the unfilled triangles show category b. Filled squares and triangles show the location on the distribution of the example pictures of the Gabor images.

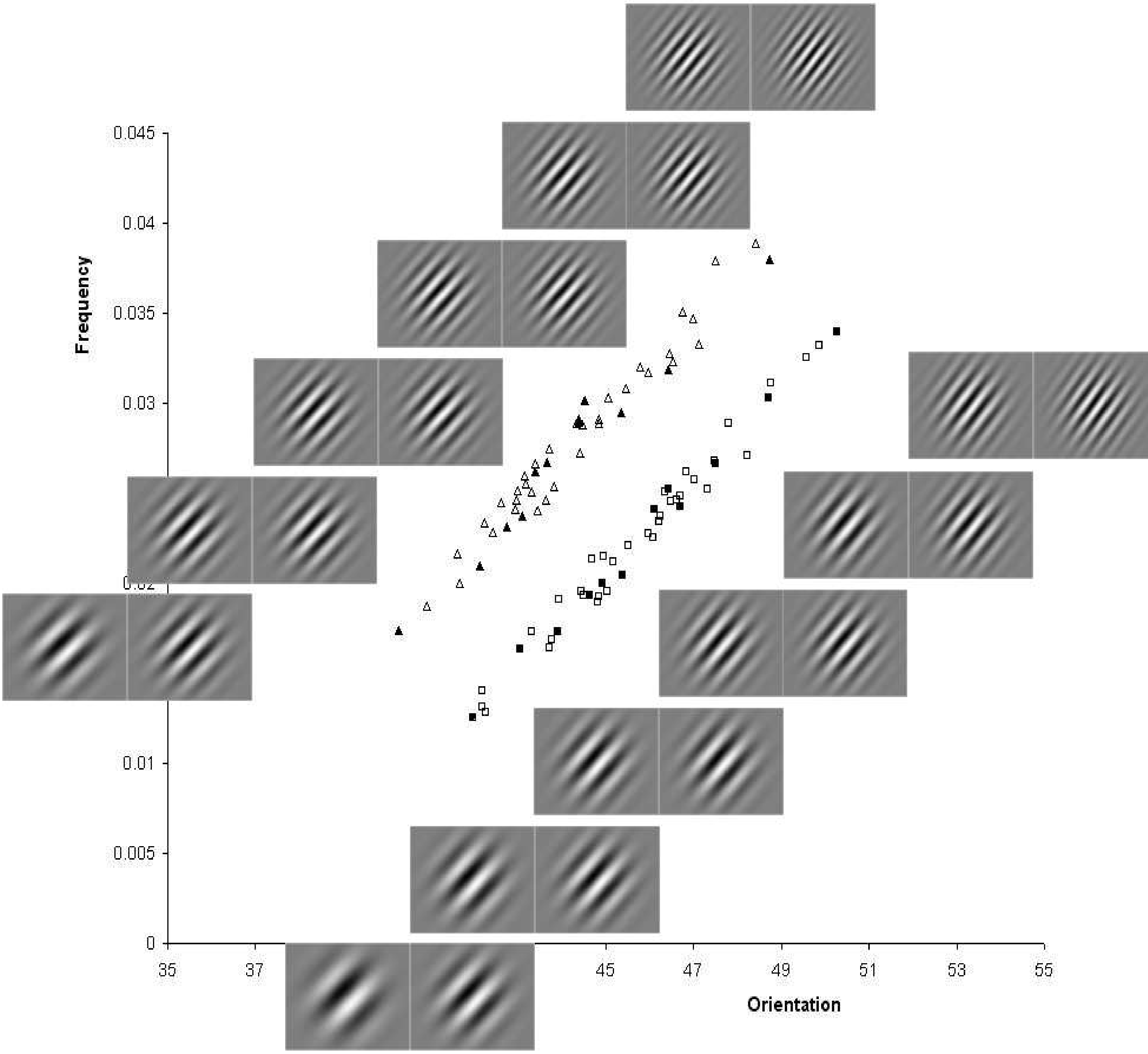


Figure 4.2. The stimuli distribution for the Narrow condition. Format and units are the same as Figure 1.

For both conditions the optimal decision bound was a diagonal line drawn through the center of the scatter plot, such that stimuli with frequencies and orientations on one side of the criterion were assigned to one category, while stimuli on the other side of the criterion were assigned to the other category. We adopted an arbitrary convention that the stimuli above and below the decision bound were designated as ‘category A’ and ‘category B’, respectively. Along with the decision bound, the scatterplot and 12 Gabor patches from each of the two categories are shown in Figures 4.1 and 4.2. Sample Gabor patches are 1/6 scale sized images from the actual images used in the sessions. The samples include the extreme values for each category (i.e., the stimuli in the lower left and upper right of the scatterplot) and also ten stimuli in between, spaced approximately equally. The sample stimuli that are pictured correlate to the filled symbols in the plots for each category. Means and standard deviations, as well as maximum and minimum values for the stimuli in each category for both of the conditions are shown in Table 4.1.

Stimuli were generated as follows: First, random numbers were sampled from a bivariate normal distribution for each of the categories, A and B. Forty-five number pairs were sampled for each category, and used to generate the frequency and orientation values for the stimuli. The distribution parameters for each category were the same as Maddox et al. (2003), and ensured that the mean frequency values were different for the categories whereas the mean orientation values were the same. The stimuli were then rotated by 45°. After the rotation the stimuli went through a linear transformation (5.98 was added to each frequency value and 245.81 added to each orientation) so that the mean and standard deviation for each of the categories matched those used by Maddox et al (2003). We refer to this type of task and stimuli as an Information Integration (II) task in keeping with the terminology that Maddox and colleagues used. The distribution parameters for both conditions are listed in Table 4.1.

	Condition			
	80°Wide		10°Narrow	
	Freq	Orient	Freq	Orient
Min	7.2559	0.0125	40.2820	0.0125
Max	87.2559	0.0389	50.2820	0.0389
SD	16.9401	0.0058	2.1175	0.0058
Mean	45.0000	0.0251	45.0000	0.0251

Table 4.1. Minimum, maximum, standard deviation and mean values of the stimuli distributions for frequency and orientation of both the Wide and Narrow conditions.

For display on the LCD screens (640 x 480 resolution), the values were converted to cycles per pixel with the following equation, where x equals the randomly generated frequency values: $(x/50+.25)/250$. The result was a frequency value in cycles per pixel that was maximized for display on a 640 x 480 screen. For the orientation dimension, numbers were converted to degrees from horizontal by first multiplying by $\pi/500$, then to degrees from horizontal by multiplying the radian value by 180 then dividing by π.

Gabor stimuli were generated in real-time using custom software written at the University of Canterbury. The algorithm used was based on the Gabor Filter, (Yao, Krolak, & Steele, 1995) and was integrated into a C++ program that displayed the images based on a pre-determined CSV file listing of frequency (cycles per pixel) and orientation (degrees).

Procedure

Because subjects were experimentally naïve, they were first shaped to peck yellow circles displayed in the center of the touch screen. They were then trained to peck the two lower right and left side keys using a modified autoshaping procedure. When subjects responded consistently both to the touch screen and keys, training began in the first condition. Sessions occurred daily and at the same time (1100h) with few exceptions. All sessions consisted of 90 trials and sessions were run until stability was achieved in each phase of each condition.

The sequence of events on experimental trials was as follows. After a 9-s inter-trial interval (ITI) during which the chamber was dark, the houselight was illuminated. One second later, the trial began with the display of a Gabor image on the touch screen. The image was maximum possible size that could be shown (640 x 480 pixels) and measured approximately 95 cm high by 125 cm wide. After pigeons had made five responses to the image the screen was darkened and the two lower keys were illuminated (e.g., left key red, right key green), signaling the choice phase. A single response to the correct key produced 3-s access to grain. During reinforcement, all illumination in the chamber was extinguished except for the feeder light. A correction procedure was used such that if the response was incorrect, the houselight flashed off and on for 10 s (1 s off, 1 s on), and the trial was repeated with the same Gabor stimulus. After five responses had been made to the screen, only the correct side key was lit and a single response produced 1.5-s access to grain.

Pigeons were exposed to the Wide 80° and Narrow 10° conditions in counterbalanced order, followed by a replication of the initial condition. Training continued in each condition until a visual stability criterion was reached. In the first condition, extended training was given because we wanted to assess the long-term stability of responding given the novel nature of the procedure. The order of conditions for each pigeon is shown in Table 4.2.

	Stimuli		
Subject	Condition 1	Condition 2	Condition 3
G6	Narrow	Wide	Narrow
G7	Wide	Narrow	Wide
G8	Wide	Narrow	Wide

Table 4.2. Order of the conditions for each of the pigeons. Wide represents the 80° condition and Narrow° the 10° condition.

RESULTS

The percentage of correct responses was calculated over the last 10 sessions in each condition. Average accuracy and standard deviations for each pigeon in both conditions are reported in Table 4.3.

	Condition							
Subject	Wide	StDev	Narrow	StDev	Wide 2	StDev	Narrow 2	StDev
G6	0.90	0.029	0.64	0.048	-	-	0.66	0.024
G7	0.93	0.028	0.64	0.047	0.89	0.039	-	-
G8	0.85	0.042	0.65	0.057	0.90	0.080	-	-
Overall	0.89	0.033	0.64	0.051	0.89	0.059	0.66	0.024

Table 4.3. Overall means and standard deviation are displayed for each subject and each condition.

In the Wide condition, accuracy was high overall for all birds (M = 89%). By contrast, accuracy was lower in the Narrow condition (M = 64%) for each pigeon. The difference was statistically significant, $t(2) = 9.45$, $p < .01$. Table 2 also shows that results were similar in the replication condition.

To show acquisition of performance, accuracy values for all subjects and sessions are plotted in Figure 4.3.

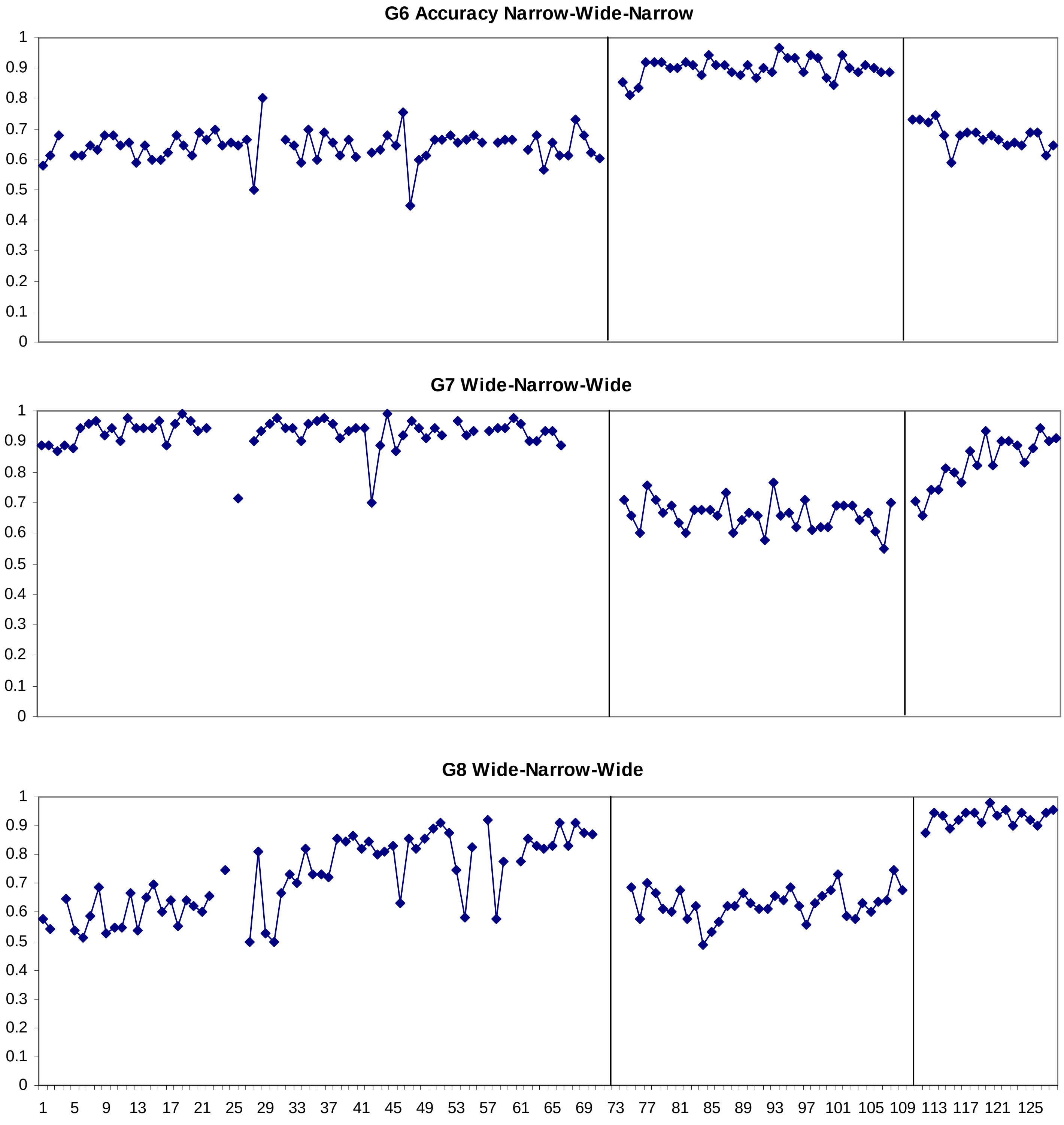


Figure 4.3. Accuracy for each session for the Wide, Narrow and Wide replication condition are reported. The session number and percent correct are shown on the x and y axis.

The upper panel shows results for G6. In the first condition (Narrow), accuracy remained approximately stable at 60% despite the extended training. When switched to the Wide condition, accuracy jumped to over 80%, evident from the very first session, and with little change over the rest of the condition. When G6 was returned to the Narrow condition, accuracy dropped to 70% and decreased somewhat over successive sessions to approximately the level from the first condition. The middle panel shows results for G7, and notably includes the only session with perfect performance (100% accuracy) in the Wide condition (session #45). When switched to the Narrow discrimination, accuracy dropped to 65% and did not change systematically across the condition. On return to the Wide condition, accuracy increased gradually to approximately the same level achieved in the first condition. Results for G8 (bottom panel) show a similar pattern as G7, but with somewhat more variability in the Wide condition. It is notable that for all three pigeons, accuracy changed abruptly at the start of the second condition and did not change systematically with continued training.

Model Analysis

In order to determine whether pigeons' responding in the Wide and Narrow conditions indicated control by a single stimulus dimension or by multiple dimensions, we conducted a series of modeling analyses. In these analyses, we compared fits of unidimensional and multidimensional models of category learning (Ashby, 1992) to individual-subject data from the last 10 sessions (900 trials) of each condition. We first provide a brief description of each of the models.

Multidimensional model (General Linear Classifier)

According to the General Linear Classifier (GLC), which is one of a family of models known as General Recognition Theory (GRT; (Ashby & Waldron, 1999); (Ashby, 1988); (Ashby & Townsend, 1986)), stimuli are represented in a two-dimensional perceptual space, similar to Figures 1 and 2. The subject learns to associate different regions of the perceptual space with different responses. The two regions in the perceptual space are defined by a linear 'decision bound'. When a stimulus is presented on a given trial, the distance of the stimulus from the decision bound determines the probability of a choice response. Specifically, the decision bound is defined as:

$$\delta X + \gamma Y + e = 0 \tag{1}$$

where X and Y are orientation and frequency, respectively, and δ, γ, and e are constants. When a stimulus X_0, Y_0 is presented on a trial, the distance of the stimulus from the decision bound is given by:

$$h(X_0, Y_0) = \frac{\delta X_0 + \gamma Y_0 + e}{\sqrt{\delta^2 + \gamma^2}} \tag{2}$$

For $h = 0$, the probability of responding category A, $P(A) = 0.50$. For $h > 0$, $P(A) > 0.50$ and for $h < 0$, $P(A) < 0.50$. Specifically, $P(A)$ is defined as the cumulative normal distribution function (Φ) evaluated at $h(X_0,Y_0)$:

$$P(A) = \Phi\left(\frac{h(X_0,Y_0)}{\sqrt{\sigma_h^2 + \sigma_c^2}}\right). \tag{3}$$

The denominator of Equation 3 represents the noise or error variance in the model, and includes terms for both perceptual (σ_c^2) and criterial variance (σ_c^2). Although other models within the GRT family can distinguish between perceptual and criterial variance (see Ashby, 1992), for the GLC only a single noise parameter, σ, is estimated which represents combined perceptual and criterial variance. Effectively, the GLC represents a generalization of signal detection theory to the two-dimensional case (Ashby & Townsend, 1986).

Thus in applying the GLC to data from the present experiment, three parameters must be estimated: the slope and intercept of the decision bound, and the noise parameter, σ. Note that the slope and intercept are defined as $-\delta/\gamma$ and $-e/\gamma$, respectively, by Equation 1.

Unidimensional models

Two unidimensional models were also considered. According to the UNI-O model, subjects respond on the basis of orientation, but variation in frequency has no effect. The UNI-F model is similar except that decisions are based entirely on frequency. Both models have two parameters: a critical value on the particular dimension (*Xcrit*) and a noise parameter, σ. For stimulus *X* presented on a given trial, the probability of responding category A is defined as

$$P(A) = \Phi\left(\frac{X - Xcrit}{\sigma}\right) \tag{3}$$

Parameter estimation

Maximum likelihood estimation was used to obtain parameters for the GLC and unidimensional models for individual-subject data. Specifically, parameter values that minimized the negative log-likelihood function were obtained through a two-step process. First, a simulated annealing algorithm (Goffe, Ferrier, & Rogers, 1994) was used to estimate a local minimum, and then parameter estimates were refined using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method (Avriel, 2003). Initial parameter values were randomly determined. Model predictions and optimization procedures were implemented in a computer program using routines in the open-source TPMATH library and compiled with Free Pascal version 2.0.2 (retrieved on 27 August 2006 from http://www.unilim.fr/pages_perso/jean.debord/tpmath/tpmath.htm and http://www.freepascal.org, respectively). Repeated simulations showed that parameter estimates were stable for all subjects and conditions and did not depend on initial values.

Model Results

Table 4 contains the VAC and AIC results for each pigeon and condition for each of the three models. Overall, the GLC provided a good quantitative description of the data, accounting for an average of 93% and 84% of the variance in the Wide and Narrow conditions, respectively. The lowest AIC values are shown in bold. For each pigeon, the GLC was the best fitting model (i.e., lowest AIC) in the Wide condition. This result was also obtained in the replication (Wide 2) conditions. By contrast, the UniF model was the best fitting model for all pigeons in the Narrow condition. These results suggest that pigeons' responding was controlled by both stimulus dimensions in the Wide condition, but only by frequency in the Narrow condition. That is, when the range of orientation values was restricted in the Narrow condition, responding was controlled exclusively by frequency.

	GLC		UniF		UniO	
	VAC	AIC	VAC	AIC	VAC	AIC
G6 Narrow	0.87	84.33	0.86	**80.36**	0.31	106.80
G6 Wide	0.94	**64.51**	-1.46	2704.01	0.00	128.76
G6 Narrow 2	0.89	78.86	0.89	**75.36**	0.28	112.51
G7 Wide	0.97	**48.71**	-1.08	2463.51	-0.01	128.77
G7 Narrow	0.90	73.73	0.90	**69.86**	0.25	107.51
G7 Wide 2	0.95	**47.23**	-1.08	2512.44	-0.01	128.77
G8 Wide	0.88	**77.90**	-1.26	2268.13	-0.05	128.77
G8 Narrow	0.71	94.60	0.68	**91.80**	0.18	108.97
G8 Wide 2	0.92	**71.34**	-1.43	2572.30	0.00	128.76

Table 4.4. Akaike information criteria (AIC) and variance accounted for (VAC) values are displayed for each of the pigeons in each condition for each of the 3 models tested. General linear classifier (GLC, unidimensional frequency (UniF) and unidimensional orientation (UniO) models were tested.

	GLC			UNI-F		UNI-O	
	Slope	Intercept	Noise	Mean	Sigma	Mean	Sigma
G6 Narrow	0.0001	0.0238	0.0049	0.0277	0.0050	46.4475	3.6685
G6 Wide	0.0004	0.0080	0.0037	0.0326	0.0151	44.9419	32.9194
G6 Narrow 2	0.0003	0.0144	0.0037	0.0326	0.0040	45.7039	3.8074
G7 Wide	0.0004	0.0100	0.0029	0.0326	0.0118	40.6836	33.3726
G7 Narrow	0.0001	0.0167	0.0037	0.0230	0.0038	43.6687	3.4951
G7 Wide 2	0.0004	0.0097	0.0029	0.0326	0.0118	41.5674	33.1554
G8 Wide	0.0003	0.0122	0.0040	0.0326	-0.0458	34.3295	49.5633
G8 Narrow	0.0006	0.0008	0.0062	0.0301	0.0075	48.6232	6.6040
G8 Wide 2	0.0003	0.0119	0.0036	0.0326	0.0107	41.7212	52.5061

Table 4.5. Model parameters for each of the models tested indicated by slope, intercept and noise.

Parameter values for each model are listed in Table 4.5. Overall, GLC parameter values were reasonably consistent across subjects for each of the conditions.

Estimates of noise parameters were also similar for the Wide and Narrow conditions.

Asymptotic Performance, Wide Condition

To examine asymptotic performance more closely and the effectiveness of the GLC in describing the data more closely, we plotted P(A) as a function of the orientation values, and also included the predictions of the GLC. These are displayed in Figure 4.4.

The left column of Figure 4 shows the plots for G6, G7 and G8 with category A on top represented by the filled squares and category B by the unfilled triangles. For all subjects, a distinct inverted U shape pattern as a function of orientation was apparent in responding to Category A stimuli. Specifically, P(A) values were highest for orientation values in the middle of the range, and decreased for orientations at both the low and high end of the range. By contrast, no systematic trend was evident for responding to category B stimuli.

The right column of Figure 4 displays the same obtained values but also includes the P(A) values predicted by the GLC for each category. These are represented by x’s for predicted category A values and +’s for category B. Here we see that the GLC does not predict well the inverted-U pattern in the obtained results. In contrast, the GLC predictions can be described as two approximately parallel linear functions, and thus do not vary systematically with orientation.

To examine deviations from the GLC predictions more closely, standardized residuals were computed. Figure 4.5 shows the standardized residuals as a function of orientation for individual pigeons for the Wide condition. For all pigeons, an inverted-U shaped pattern was obtained for category A residuals (filled squares), whereas no systematic trend was apparent for category B residuals (unfilled triangles). There was an increasing linear trend in the category B residuals for G6, but not the other pigeons. This trend resulted because the GLC predicted a decreasing linear trend for category B for this pigeon (parallel with the category A predictions), but performance did not vary systematically with orientation.

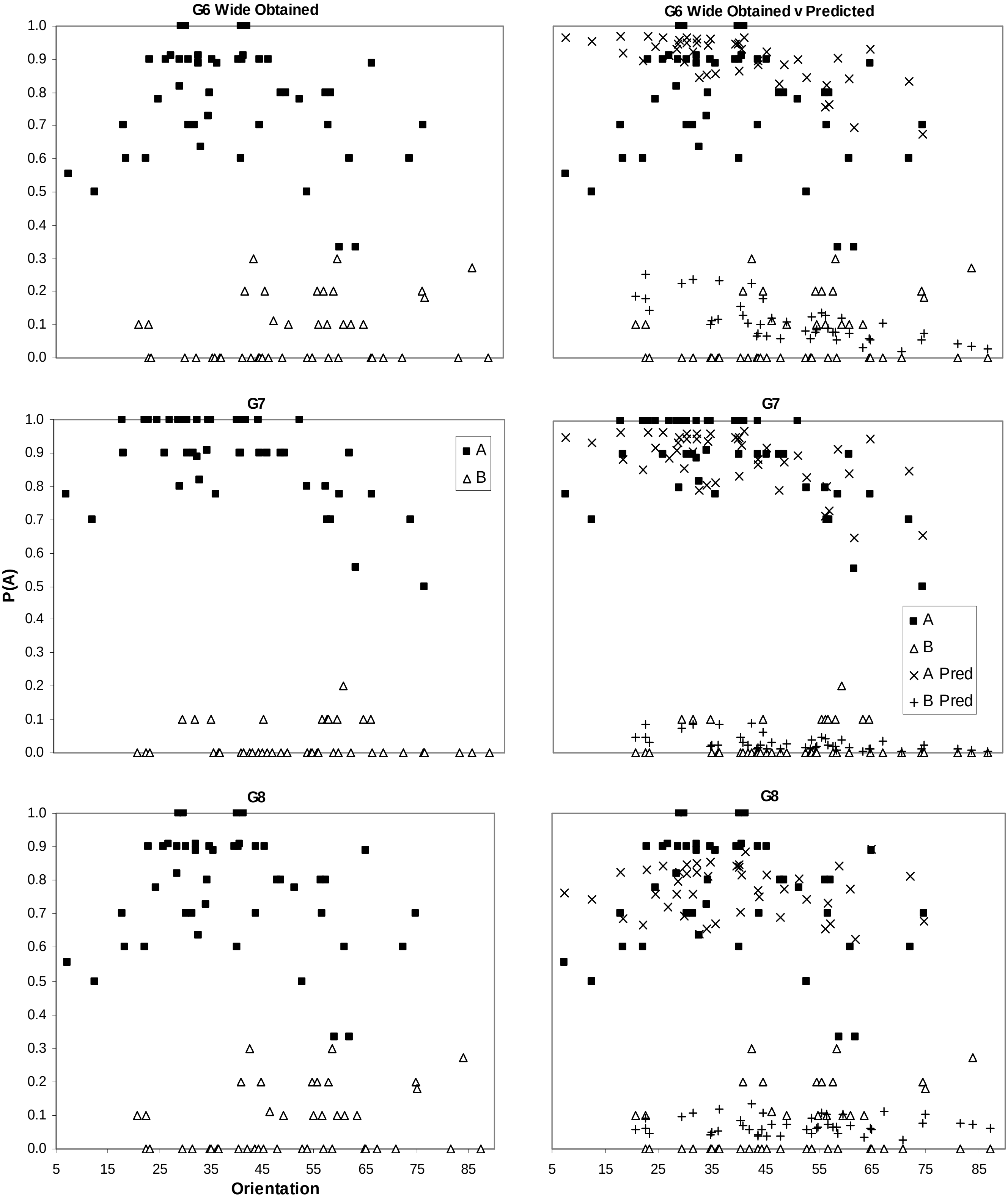


Figure 4.4. Probability of a category A response (P(A) in the y axis plotted against orientation in the x axis. The wide condition obtained data is displayed for all 4 subjects in the left column and the right contains obtained and predicted values for comparison. Category is A represented by the filled squares and b the unfilled triangles.

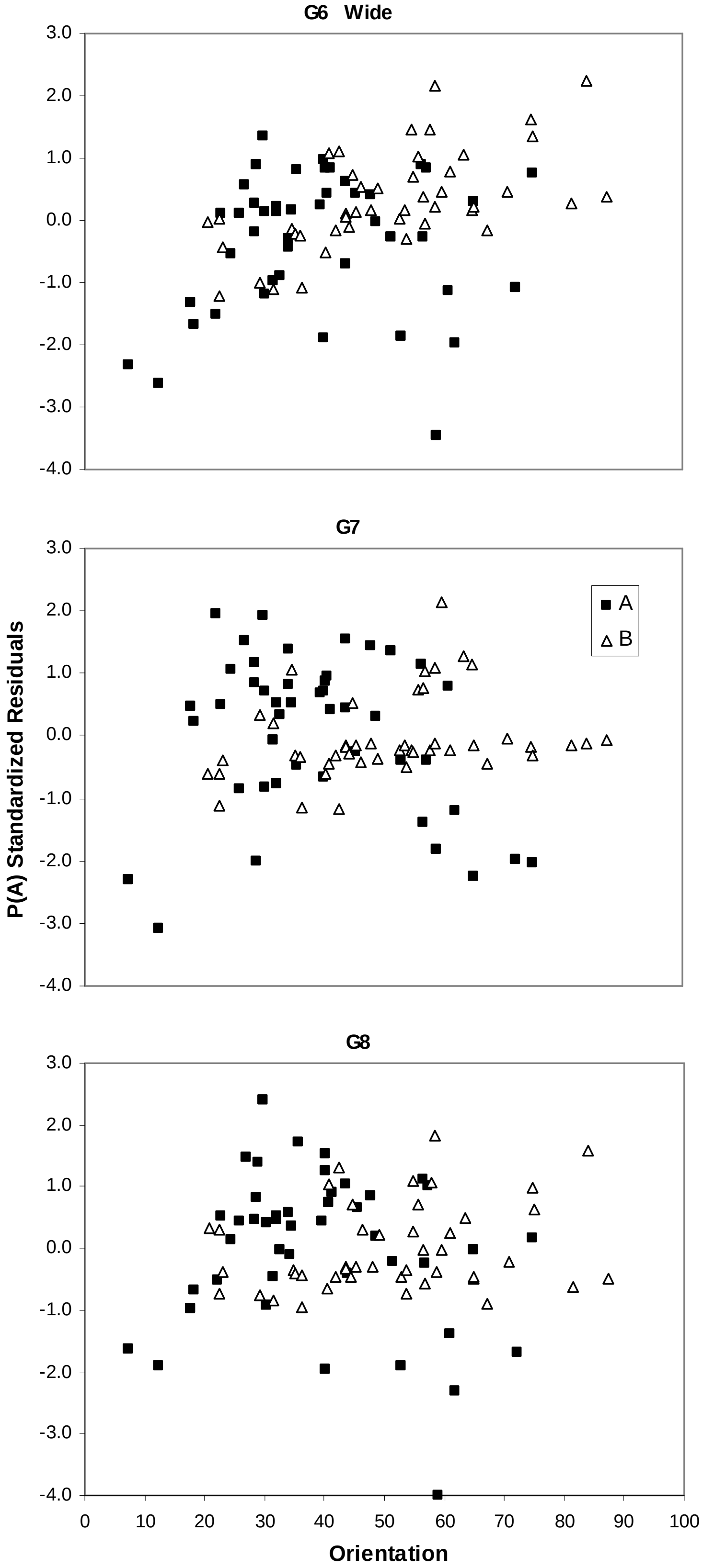


Figure 4.5. Standardized residuals from the GLC fits for each subject.

Figure 4.6 provides a summary of the GLC fits to the Wide condition data based on the average results. The top panel of Figure 4.6 confirms the inverted U shape for category A stimuli, whereas accuracy is approximately constant across the range of orientations for category B stimuli. The second (middle) panel plots the same average obtained P(A) data as the top panel with the addition of values predicted by the GLC. The middle and bottom panels show the predicted and obtained values, and the standardized residuals, respectively. The standardized residuals for Category A show an inverted U shaped pattern, confirming that the GLC is unable to account for the Category A results in the top panel. However, the residuals for Category B show a slight increasing linear trend. Although P(A) for Category B in the upper panel did not show such a trend, the GLC predictions for the two categories are constrained to have the same slope for orientation (middle panel). That is, because predicted P(A) decreases as a function of orientation for Category A – likely caused by the greater decrease in obtained P(A) for the stimuli with higher orientation values – predicted values for Category B also decrease. These decreased because the only way for the GLC to predict that accuracy decreases more for high orientation Category A stimuli is by rotating the decision bound (in this case, towards the high orientation Category A stimuli). As a result, the decision bound moves further away from the high orientation Category B stimuli, resulting in increased accuracy (and lower P(A) values).

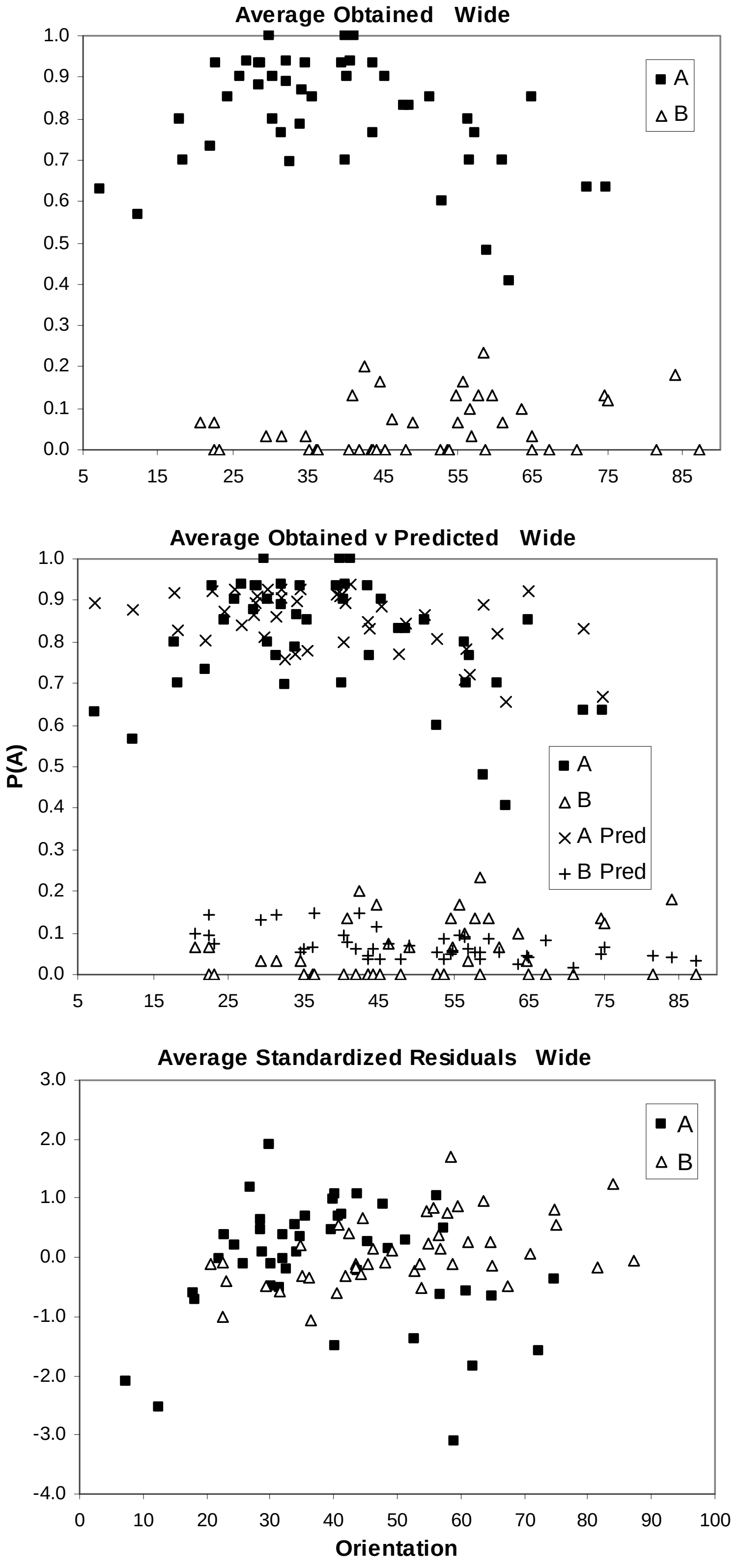


Figure 4.6. For the Wide condition the top panel shows the overall average obtained values. The middle panel shows the average obtained vs. predicted values and the bottom panel displays the average standardized residuals from the GLC model fits.

Asymptotic Performance, Narrow Condition

Comparable asymptotic results for the Narrow condition are shown in Figure 4.7, which shows the P(A) values as a function of orientation (left column) and with the predictions of the GLC when fitted to the individual data (right column). For all subjects, P(A) increased with orientation for both categories. However, recall that the model fits (Table 4.3) showed that responding was controlled by frequency but not orientation, so this increase can be attributed to the fact that orientation and frequency were positively correlated (Figure 4.2). GLC predictions are added in the right-hand column, and show that the model provided a good account of the results, with no systematic deviations apparent.

Standardized residuals from GLC fits to the Narrow condition data are shown in Figure 4.8. Across subjects, there was no evidence of a systematic pattern in the residuals. Together with the reasonably high percentage of variance accounted for by the model (84% on average), suggests that the GLC provided a good account of responding in the Narrow condition. Figure 4.9 provides a summary of the obtained data from the Narrow condition (upper panel), and the GLC predictions (middle panel) and residuals (lower panel) in terms of the average results.

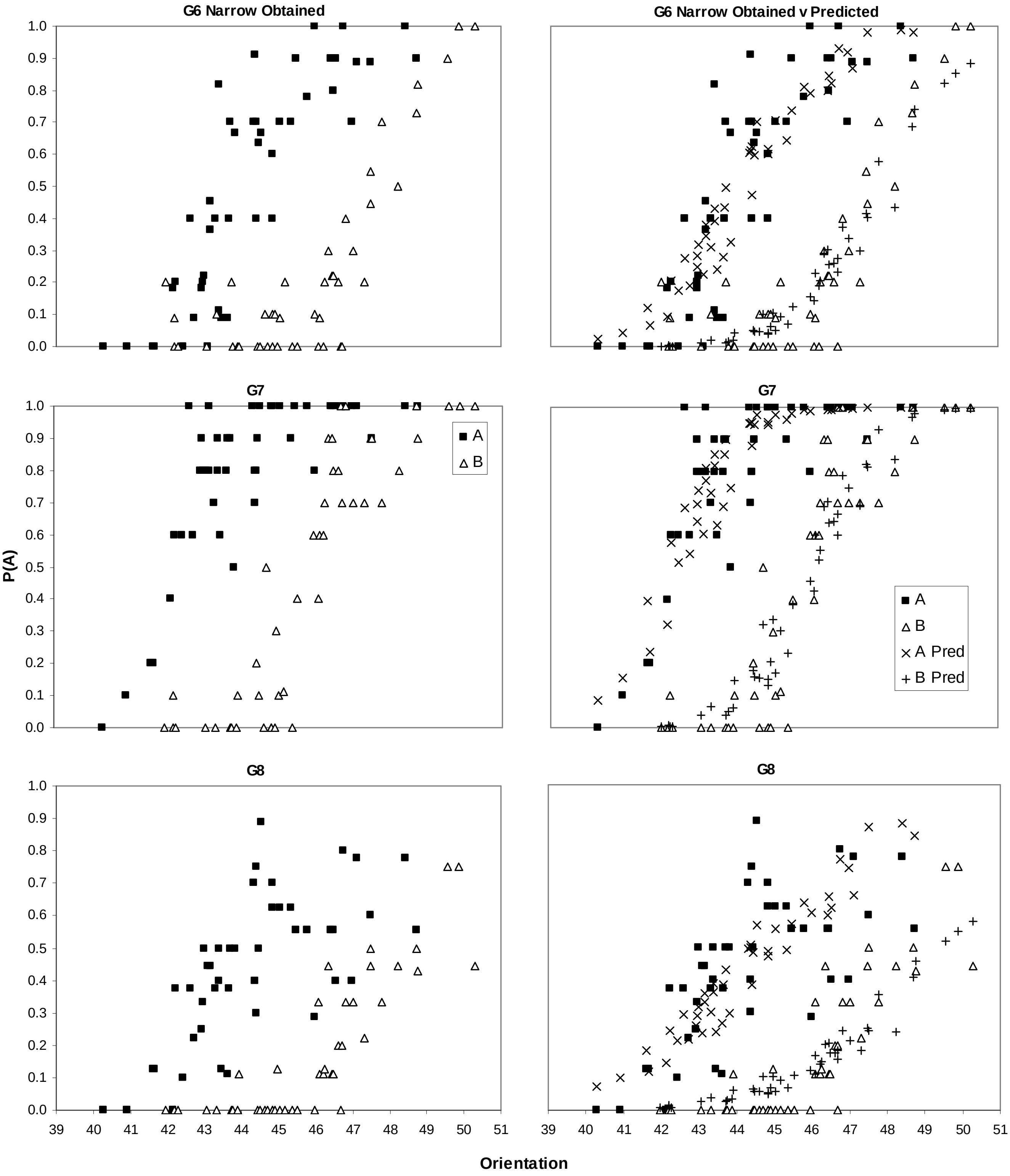


Figure 4.7. Probability of a category A response (P(A) in the y axis plotted against orientation in the x axis. The narrow condition obtained data is displayed for all 4 subjects in the left column and the right contains obtained and predicted values for comparison. Category is A represented by the filled squares and b the unfilled triangles.

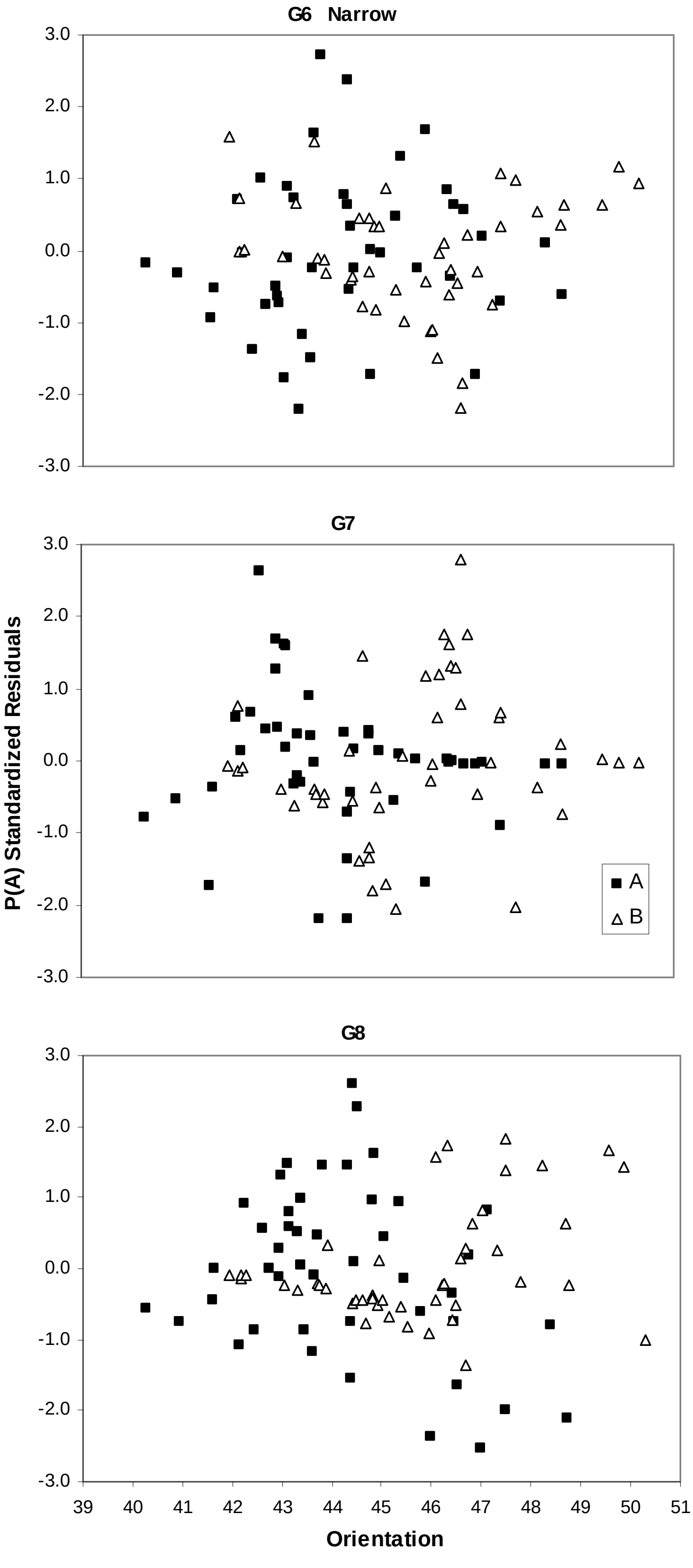
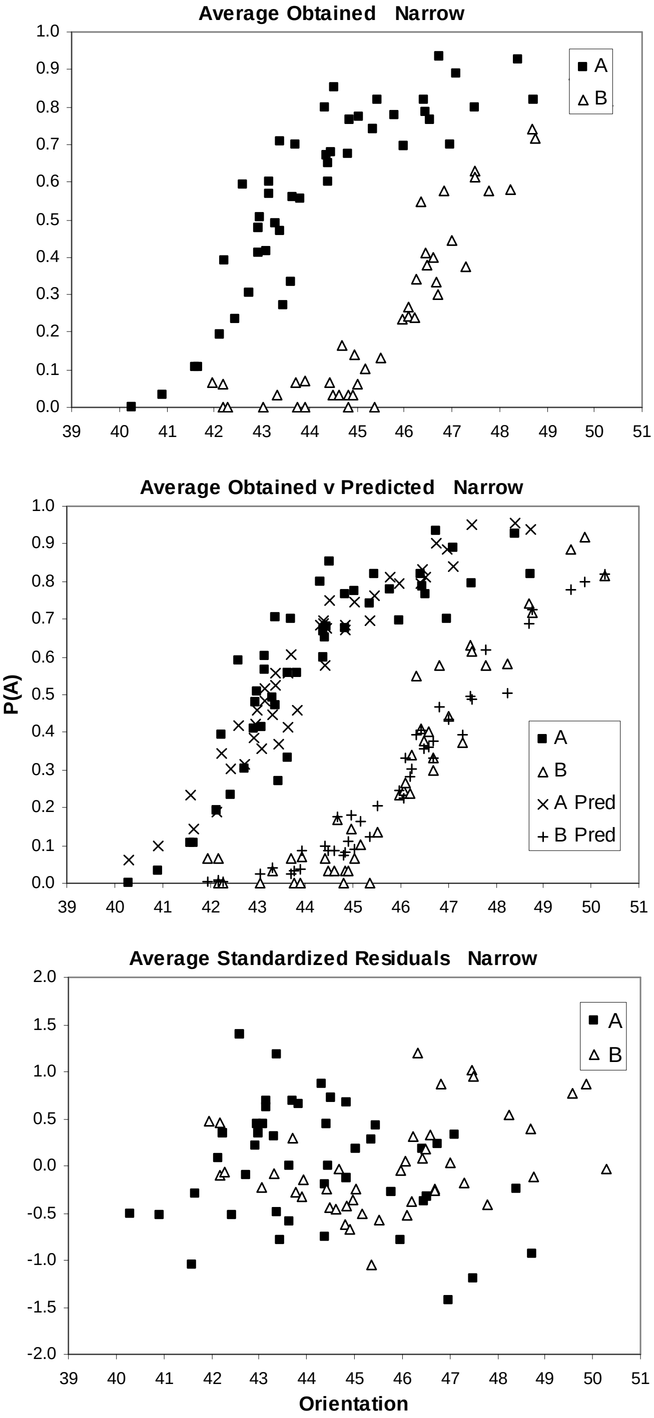


Figure 4.8. Standardized residuals from the GLC fits for each subject.



Average Obtained Narrow

Average Obtained v Predicted Narrow

Average Standardized Residuals Narrow

Figure 4.9. For the Narrow condition the top panel shows the overall average obtained values. The middle panel shows the average obtained vs. predicted values and the bottom panel displays the average standardized residuals from the GLC model fits.

Finally, we conducted a series of polynomial regression analyses to determine whether any systematic trends in the residuals were present for both Wide and Narrow conditions. For these analyses, the standardized residual values were regressed on the orientation (centered) and orientation squared values for individual as well as average data. In this way, both linear and quadratic trends could be identified.

Pigeons	**Category A**			**Category B**		
Wide	β Linear	β Quadratic	R^2	β Linear	β Quadratic	R^2
G6	0.25	-0.43 **	0.19 *	0.61 **	-0.12	0.37 ***
G7	-0.02	-0.62 ***	0.39 ***	0.31 *	-0.25	0.14 *
G8	-0.03	-0.45 **	0.21 **	0.24	-0.05	0.06
Mean	0.07	-0.59 ***	0.33 **	0.51 ***	-0.17	0.26 **
Narrow						
G6	0.13	-0.20	0.04	-0.05	0.56 ***	0.31 ***
G7	-0.12	-0.10	0.03	0.19	-0.09	0.04
G8	-0.17	-0.44	0.26 **	0.36 *	0.13	0.16 *
Mean	-0.09	-0.45 **	0.23 **	0.30 *	0.31 *	0.21 **

Table 4.6. The left most column shows the label for each pigeon followed by the mean and the bold text divides the conditions, wide and narrow. Category A is on the left set of results columns and category B on the right. Each category reports the Beta Weights (β) for the linear coefficient, β for the quadratic coefficients and the R^2 values. Significant regression results for β Linear or β Quadratic are indicated with an asterisk (*) for 0.1, 0.05, 0.01 significance levels using *, ** and *** respectively.

Results are shown in Table 4.6. The left column shows the pigeon name and condition label (top half = wide condition, bottom half = narrow condition) and the left three columns of data correspond to category A and the right side category B data. The data reported are beta weights (β) for the linear and quadratic functions as well as overall R^2 values. In category A, wide condition, the quadratic function is negative and significant for all three subjects as well as the average. This result confirms the inverted-U pattern that was identified in the earlier plots. The R^2 values were also significant indicating an overall significant regression result for all three subjects. For Category B, a significant positive linear component was obtained for two of three pigeons, as well as for the average data, confirming the trend noted in Figure 4.6. By contrast, results for the Narrow condition show no systematic trend across subjects. Although several components were significant, these were isolated and not obtained for more than one subject. Thus, results indicate that the residuals from the GLC fits to the Wide condition showed a systematic pattern across subjects, whereas those from the Narrow condition did not.

EXPERIMENT 2

METHOD

Subjects

Four humans (2 females, 2 males), designated C, M, S and T participated as subjects and were compensated for their time with chocolate. Subjects were chosen from the local community on a volunteer basis and varied in age from 21 to 35 years old. All subjects had 20/20 vision or wore glasses that corrected their vision to 20/20. Subjects were randomly assigned to the conditions in the experiment. All were experimentally naïve with no previous experience in a psychological study.

Apparatus

Experimental sessions were conducted in a quiet, dimly lit room. A laptop was used to control the experiment, display the stimuli and record the responses. The LCD screen on the laptop was an 8.9" 1024x600 Active Matrix display but resolution during the experiment was set to 640 x 480 so as to match the resolution of Experiment 2. The A and L keys of a standard QWERTY style keyboard were used as response keys for categorization of the on-screen images. The space bar was used to indicate readiness to begin the experiment.

Stimuli

The stimuli used in experiment two were identical to those used in experiment one. Table 1 shows the parameters of the stimuli distribution and Figures 1 & 2 show stimuli examples and distributions for each of the two conditions (Wide 80° and Narrow 10°).

Procedure

Subjects were seated in a quiet, dimly lit room and instructed that the session could last up to one hour, and that accuracy not speed was the important factor in the experiment. The subjects were then asked to read the following on-screen text:

> "Thank you for your participation in this experiment on visual perception. Press the space bar to view the instructions and begin."

Upon pressing the space bar on the keyboard the next screen read:

> "This experiment will last approximately 50 minutes, with 3 brief rest periods. When you begin an image will be displayed on screen. After viewing it you will be asked to categorize the image in either category 1 by pressing the A key, or 2 by pressing the L key. Both categories of images are displayed randomly and are equally likely to be displayed. Do your best to learn about the categories as perfect performance in possible. Speed in not important only accuracy so categorize the images as accurately as you can. A prize will be awarded to the person with the best performance. You will now be given 20 practice trials. Press the space bar when you are ready to begin."

Subjects were then given 20 practice trials. The stimuli displayed during the practice trials were prototype stimuli for each of the two categories. The two prototypes were displayed in random order, one for each of the categories, and were based on the mean frequency and orientation values for each category. All trials were response terminated and subjects were allowed to view the stimuli for as long as they wished. After completion of the practice trials, subjects were asked if they had any questions about the experiment (none of the four subjects had any questions), and the experimenter left the room. When the subject pressed the space bar, the regular session began with four blocks of 90 trials and rest periods in-between each block. Stimuli were independently randomized for each block and between each of the two sessions in each condition. Subjects were given feedback on-screen after pressing the A or L categorization keys which indicated whether the response was “Correct” or “Incorrect”. The order of conditions, Wide (80°) and Narrow (10°) and response key assignments were counterbalanced across subjects as shown in Table 4.7.

	Condition & Key Order			
Subject	Stimuli	Cat A Key	Stimuli	Cat A Key
C	Wide	A	Narrow	L
M	Wide	L	Narrow	A
S	Narrow	A	Wide	L
T	Narrow	L	Wide	A

Table 4.7. Condition and key order for each of the subjects. Capital A or L represents the keyboard letter used for the category A assignment.

Subjects always received 20 practice trials with feedback at the start of each of the four sessions followed by the 4 blocks of 90 trials. The four sessions consisted of two sessions in each condition, either narrow or wide. Sessions were conducted in the evenings on convenient days for the subjects and took approximately two weeks for each subject to complete. Chocolate was given as incentive to participate in the sessions and after all of the sessions were completed the subject with the highest overall accuracy average received a large 500 gram chocolate bar.

RESULTS

Figure 10 shows the percentage of correct responses for individual subjects across the four blocks of the last session in each of the two conditions. Each block contained 90 stimulus presentations. Results for subjects C, M, T and S are shown in individual panels as noted, and the X axis labels indicate whether they were in the Wide 80° or Narrow 10° conditions as outlined in Table 4.7. Overall, accuracy appeared to be higher in the Wide 80° condition regardless of condition order. The percentage of correct responses is listed for each subject and condition in Table 4.8. Averaged across subjects, response accuracy was significantly higher in the 80° condition compared to the 10° condition, M's = 80% and 66%, respectively, $F(1,3) = 23.84$, $p < 0.02$.

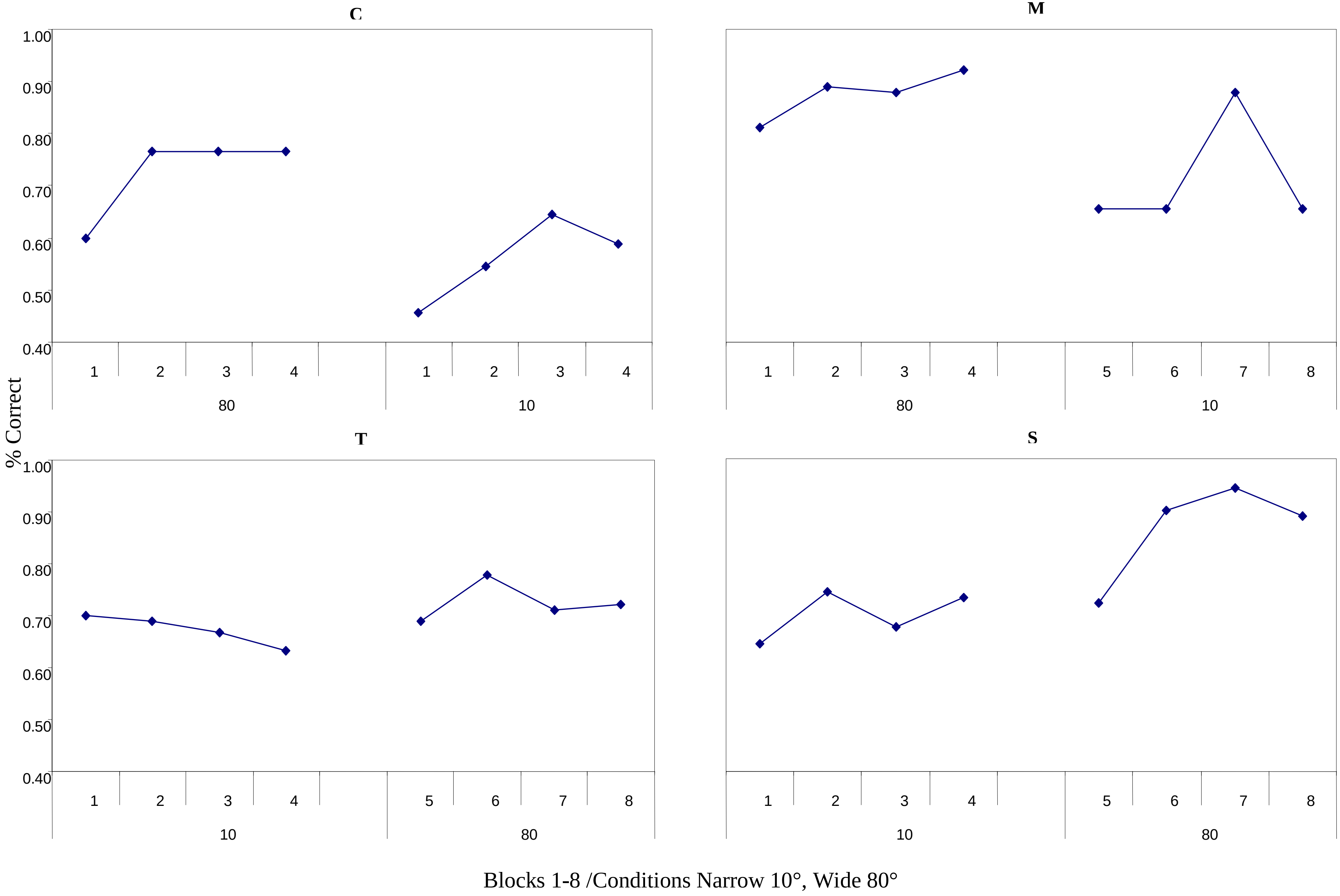


Figure 4.10. Blocks 1-4 and 5-8 represent sessions conducted in the experiment. The numbers 80 or 10 represent either the Wide or Narrow condition. Percent correct is on the x axis.

Condition	Subject	Block 1	Block 2	Block 3	Block 4	Average	StDev
Narrow	C	0.46	0.54	0.64	0.59	0.56	0.080
	M	0.66	0.66	0.88	0.66	0.71	0.111
	S	0.64	0.74	0.68	0.73	0.70	0.047
	T	0.70	0.69	0.67	0.63	0.67	0.029
Wide	C	0.60	0.77	0.77	0.77	0.73	0.083
	M	0.81	0.89	0.88	0.92	0.88	0.047
	S	0.72	0.90	0.94	0.89	0.86	0.097
	T	0.69	0.78	0.71	0.72	0.73	0.038
Significant for Condition: F(1,3) = 23.84,p<0.02							

Table 4.8. Accuracy for each subject and condition divided into 4 blocks. The overall averages and standard deviations are shown in the right hand columns. Results from a repeated measures ANOVA are displayed in the bottom cell.

Model Analyses

The GLC, UniO and UniF models were fitted to the data and compared using the same procedure as Experiment 1. For each data set, Table 4.8 indicates the best-fitting model by displaying the lowest AIC value in boldface. The variance accounted for (VAC) by each model is also listed. Table 4.9 shows that for the Wide 80° range conditions, the GLC had the lowest AIC values for all subjects. By contrast, for the Narrow condition the UniF model had the lowest AIC for all subjects. This indicates that in the Wide condition, where accuracy was significantly higher, the GLC was the best fitting model. By contrast, in the Narrow condition, in which accuracy was lower, the UniF model fit the data best. These results are similar to those obtained with pigeons in Experiment 1 and show that in the Wide condition, responding was controlled by both orientation and frequency, whereas in the Narrow condition responding was controlled solely by frequency. This suggests that when the range of stimulus orientation values was restricted, control by orientation decreased.

	GLC		UniF		UniO	
Narrow	VAC	AIC	VAC	AIC	VAC	AIC
C	0.54	112.69	0.47	**111.20**	0.06	126.41
M	0.77	87.09	0.76	**83.37**	0.25	114.65
S	0.67	96.44	0.64	**95.56**	0.12	123.03
T	0.59	106.71	0.54	**105.44**	0.08	124.25
Wide						
C	0.66	**97.89**	0.43	104.27	0.00	124.03
M	0.83	**65.74**	0.36	103.57	0.02	127.02
S	0.76	**75.80**	0.13	119.69	0.13	119.70
T	0.56	**100.02**	0.01	128.05	0.42	106.62

Table 4.9. VAC and AIC values for each subject in each condition for each of the 3 models tested.

	GLC			UNI-F		UNI-O	
	Slope	Intercept	Noise	Mean	Sigma	Mean	Sigma
C Wide	0.0002	0.0189	0.0051	0.0281	0.0084	101.0000	162.0000
C Narrow	0.0010	-0.0187	0.0071	0.0248	0.0094	44.6565	10.3648
M Wide	0.0002	0.0149	0.0030	0.0259	0.0076	36.7136	105.9640
M Narrow	0.0002	0.0160	0.0044	0.0261	0.0046	45.5734	4.0511
S Narrow	0.0008	-0.0090	0.0048	0.0253	0.0061	45.0908	6.4798
S Wide	0.0003	0.0087	0.0040	0.0238	0.0142	48.9113	41.8874
T Narrow	0.0009	-0.0142	0.0062	0.0265	0.0081	46.2934	8.6180
T Wide	0.0006	-0.0034	0.0094	0.0224	0.0562	43.0996	24.4935

Table 4.10. Model parameters for each subject and each condition for each of the three models tested.

Parameter values for each model fit to individual data are listed in Table 4.10. Overall, GLC parameter values were reasonably consistent across subjects for each of the conditions. Estimates of noise parameters were also similar for the Wide 80° and Narrow 10° conditions.

Analyses of Asymptotic Performance

To examine the model and asymptotic performance in more detail we generated plots of the orientation values and obtained probability of Category A response for each of the subjects in the last session of each of the two conditions. The plots are displayed in the left columns of Figure 4.11. In the following analyses, we first examine visually whether there was any systematic pattern in the obtained data. Because perfectly accurate responding was possible, any such pattern might indicate how performance deviated from optimality. Next we compared GLC predictions with the obtained data and noted whether there appeared to be any systematic deviations. Finally, we report results of polynomial regression analyses which tested whether the GLC residuals for individual-subject fits showed significant trends.

Asymptotic Performance – Wide Condition

The plots show the obtained probabilities of a Category A response (P(A)) separately for Category A and B stimuli, as a function of orientation. Category A stimuli are represented by the filled squares and unfilled triangles represent the Category B stimuli. Data for individual subjects in the Wide condition are shown in separate scatterplots in the left column of Figure 4.11. For Category A stimuli, the scatterplots generally show an inverted U shape that is best exemplified by subject M. The highest accuracy is achieved in the middle of the orientation range, and decreases for relatively low and high orientation values. The pattern is also visually evident in subject S's plot and, to a somewhat lesser extent, subject T. Subject C does not display the same pattern, but note that C's performance was also atypical in that their responding was overall less accurate than the other subjects.

In the right hand column of Figure 4.11, scatterplots show both the obtained values and those predicted by the GLC based on the fit to the individual data. This allows direct comparison between the obtained and predicted values and may reveal systematic deficiencies of the GLC when used to predict responding in this type of task. The deficiencies are especially evident in plots for subjects M and S. For both categories, the model’s predictions can be described roughly in terms of two parallel lines, and are unable to account for the inverted U pattern for category A.

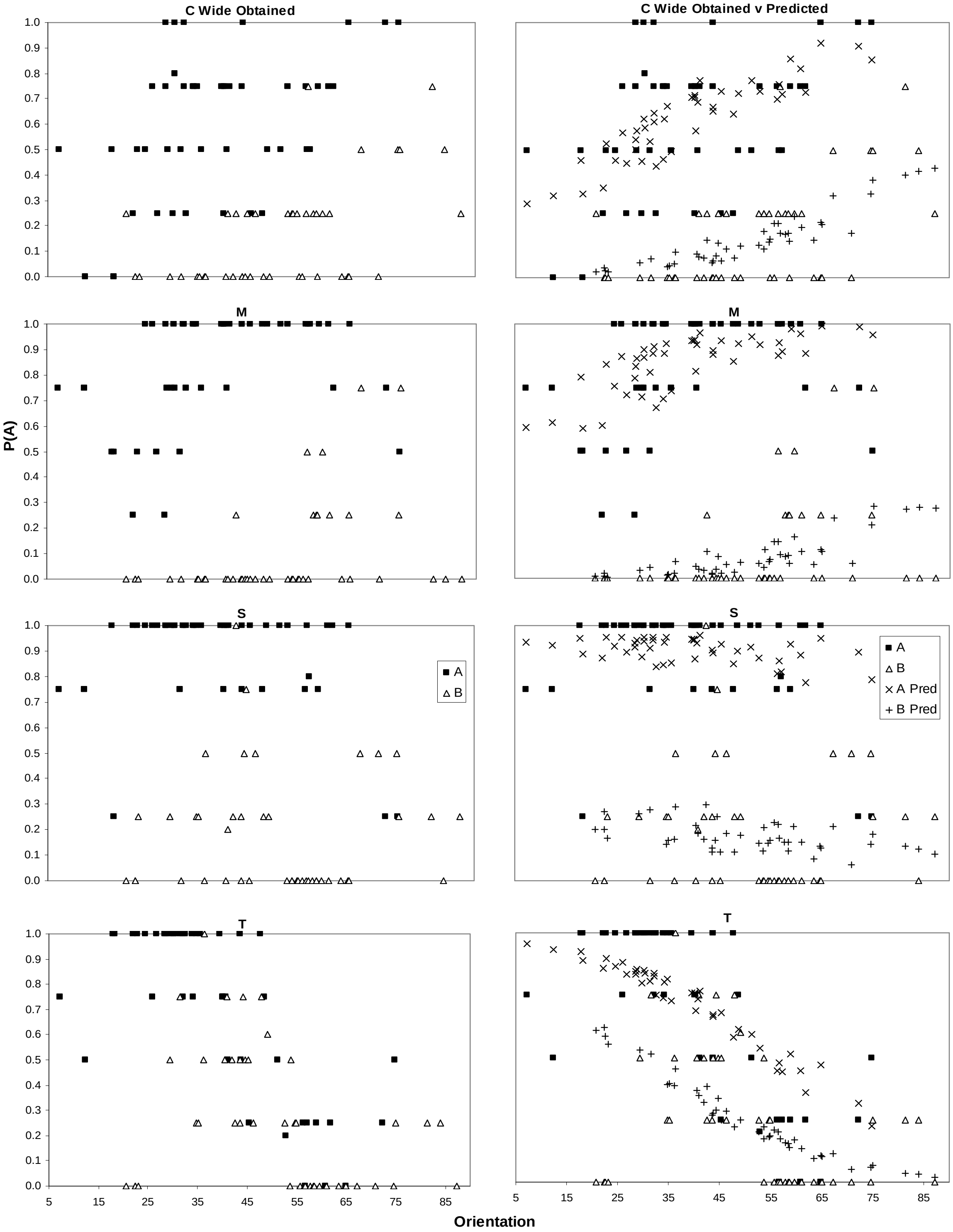


Figure 4.11. Probability of a category A response (P(A) in the y axis plotted against orientation in the x axis. The wide condition obtained data is displayed for all 4 subjects in the left column and the right contains obtained and predicted values for comparison. Category is A represented by the filled squares and Category B by the unfilled triangles.

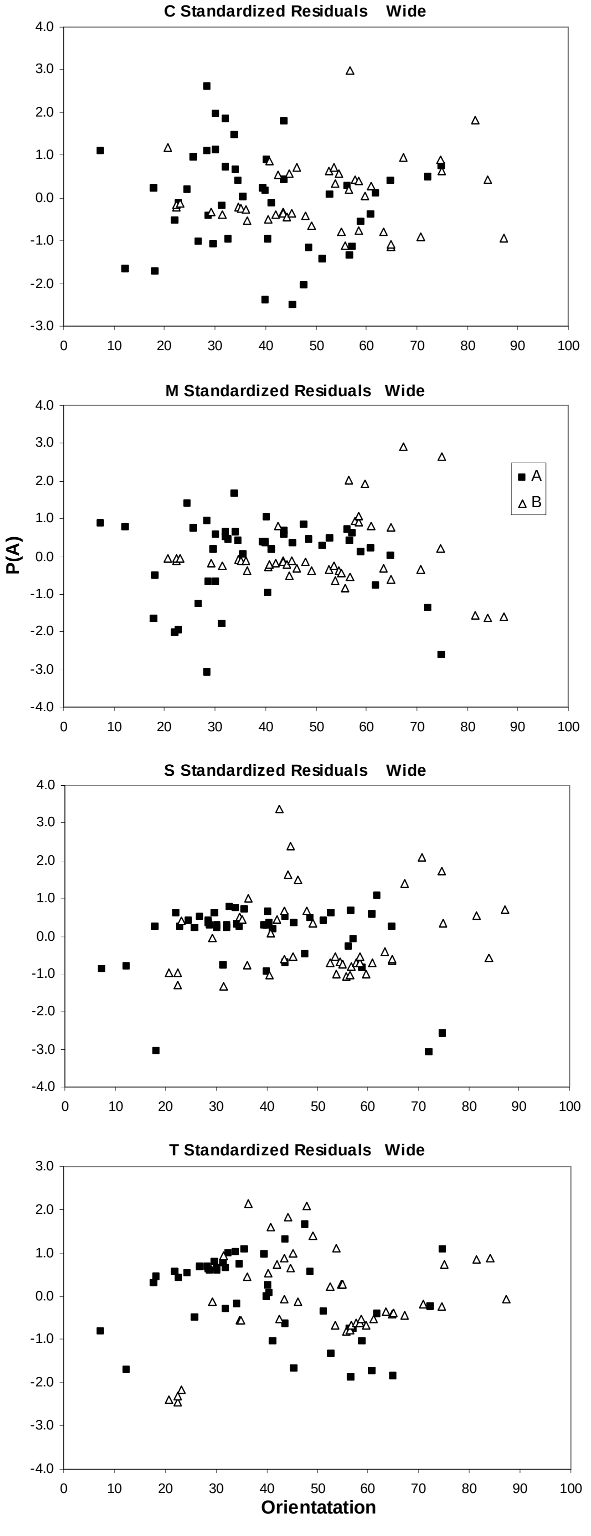


Figure 4.12. Standardized residuals from the GLC fits to the Wide condition data for each subject.

To assess the adequacy of the GLC fits to the individual data more systematically, standardized residuals by subtracting the predicted from obtained P(A) values, and dividing by the standard deviation of the obtained values. Standardized residuals for the Wide condition are shown for individual subjects in Figure 4.12, plotted as a function of stimulus orientation.

Once again we can see from the Wide condition plots that Category A responding in subjects M and S, and to a lesser extent, subject T, show evidence of an inverted U shape pattern as a function of orientation. Results for Subject C do not show any identifiable pattern, and this may be related to their lower accuracy in this condition.

To provide a summary of participants' responding in the Wide condition, Figure 6 shows the results averaged across subjects. The upper panel shows the obtained P(A) plotted against orientation, in which the inverted U shape for Category A is clearly apparent. Accuracy is highest in the middle range of and tapers off at the extreme orientation values of the category. For Category B, results are less clear but may show a modest increase as a function of orientation and perhaps an increasing linear trend. The linear trend implies that accuracy decreases as orientation increases. The middle panel of Figure 4.13 compares the same average P(A) values as the top panel with the addition of the GLC predictions for each category. This figure shows that the GLC predicts two approximately parallel linear functions, which do not correspond to the obtained data. The standardized residuals are shown in the bottom panel, which show an inverted U shape for Category A, and an increasing trend for Category B.

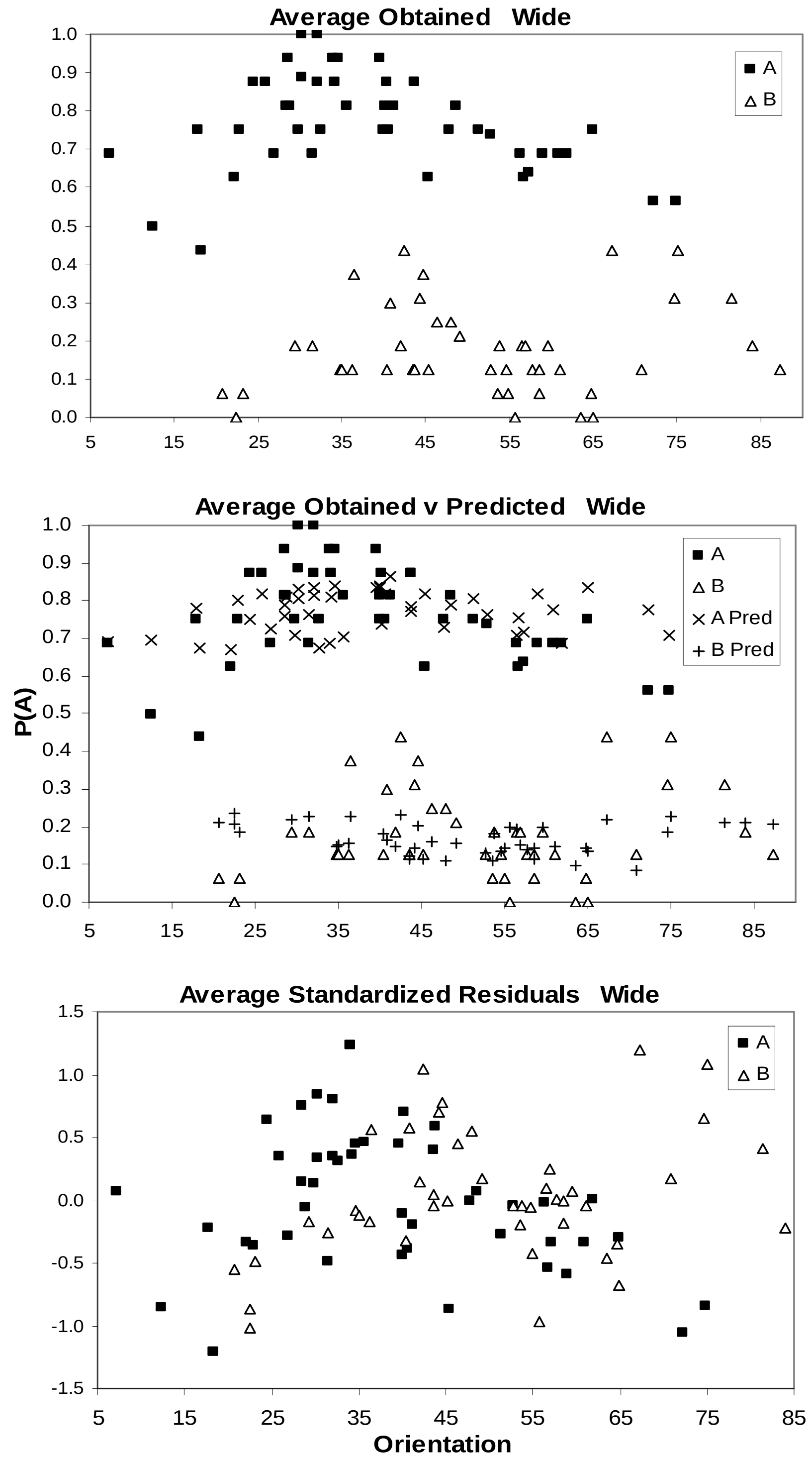


Figure 4.13. For the Wide condition the top panel shows the overall average obtained values. The middle panel shows the average obtained vs. predicted values and the bottom panel displays the average standardized residuals from the GLC model fits.

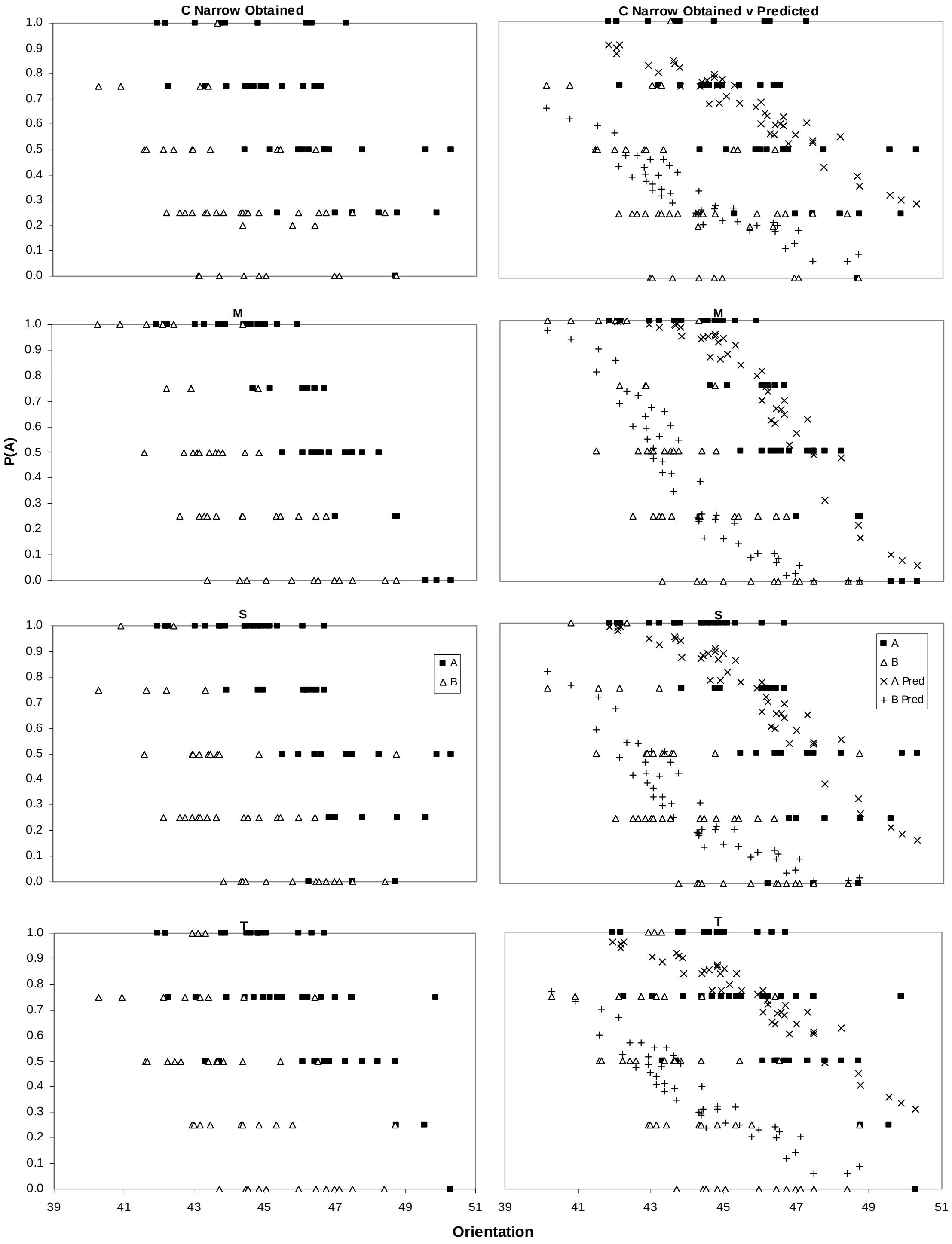


Figure 4.14. Probability of a category A response (P(A) in the y axis plotted against orientation in the x axis. The narrow condition obtained data is displayed for all 4 subjects in the left column and the right contains obtained and redicted values for comparison. Category is A represented by the filled squares and b the unfilled triangles.

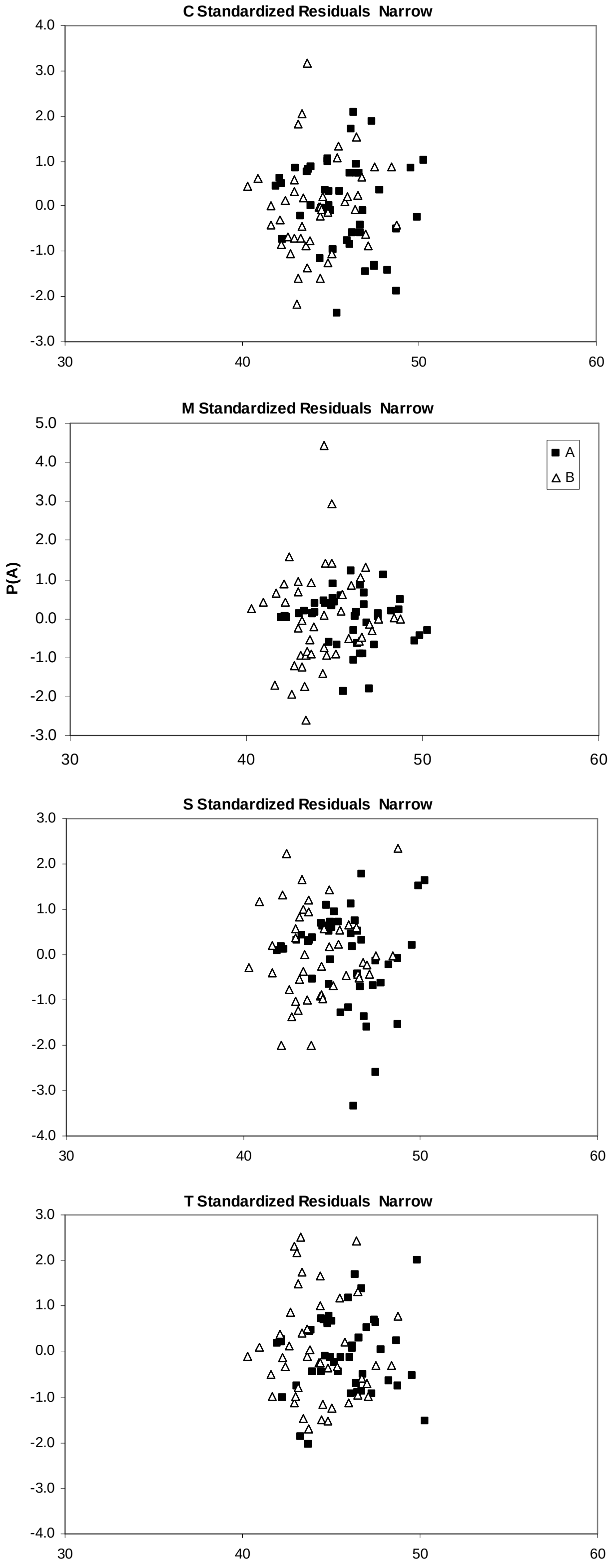


Figure 4.15. Standardized residuals from the GLC fits for each subject in the Narrow condition.

Asymptotic Performance, Narrow Condition

Figure 4.14 shows the obtained P(A) as a function of stimulus orientation for individual subjects (left column), and with the predictions of the GLC (right column). The scatterplots in the right column of Figure 4.14 show the obtained values as a function of the GLC predictions based on the fit to the individual data. The GLC was able to capture the overall trend in the data. Standardized residuals for the Narrow condition are shown for individual subjects in Figure 4.15, plotted as a function of stimulus orientation. No systematic deviations were apparent.

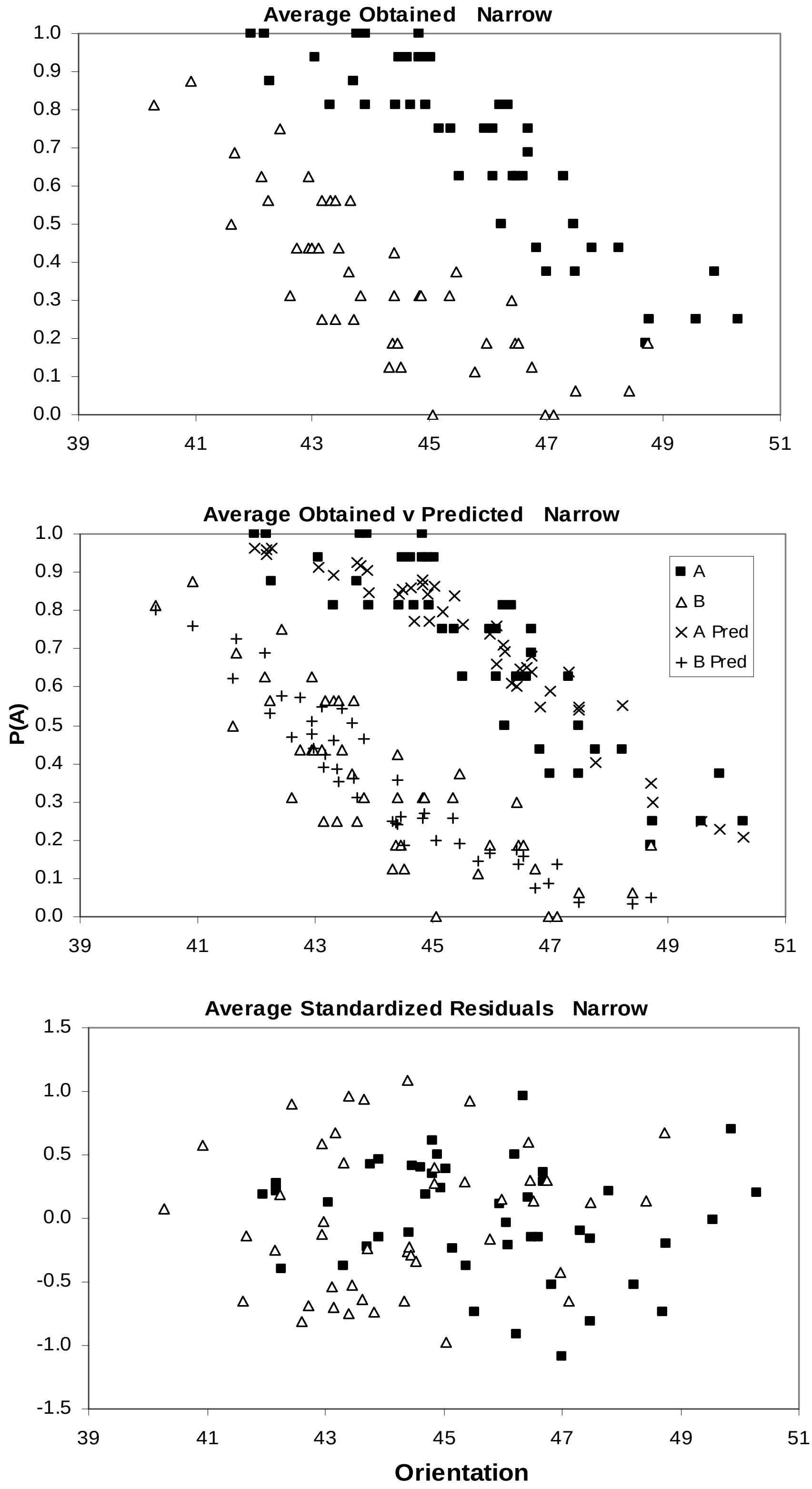


Figure 4.16. For the Narrow condition the top panel shows the overall average obtained values. The middle panel shows the average obtained vs. predicted values and the bottom panel displays the average standardized residuals from the GLC model fits.

In order to test whether systematic trends were present in the residuals from the GLC fits to the data from the Wide and Narrow conditions we conducted a series of polynomial regressions. Results are shown in Table 4.11.

Humans	Category A			Category B		
Wide	β Linear	β Quadratic	R^2	β Linear	β Quadratic	R^2
C	-0.12	0.06	0.01	0.09	0.06	0.01
M	0.07	-0.34 *	0.11 *	0.08	-0.27	0.05
S	0.02	-0.63 ***	0.39 ***	0.12	-0.01	0.01
T	-0.24	-0.23	0.14 *	0.27	-0.36 *	0.18 *
Mean	-0.13	-0.50 ***	0.31 ***	0.28	-0.29 *	0.14 *
Narrow						
C	-0.16	0.08	0.03	0.09	0.10	0.02
M	-0.15	-0.02	0.03	-0.01	0.12	0.01
S	-0.14	0.22	0.06	0.02	0.21	0.05
T	0.10	-0.16	0.03	-0.05	-0.01	0.00
Mean	-0.18	0.09	0.04	0.09	0.13	0.03

Table 4.11. The left most column shows the label for each pigeon followed by the mean and the bold text divides the conditions, wide and narrow. Category A is on the left set of results columns and category B on the right. Each category reports the Beta Weights (β) for the linear coefficient, β for the quadratic coefficients and the R^2 values. Significant regression results for β Linear or β Quadratic are indicated with an asterisk (*) for 0.1, 0.05, 0.01 significance levels using *, ** and *** respectively.

The left-most column of Table 4.11 indicates the subject and condition (Wide 80°, Narrow 10°). The left and right sides of the table show beta weights (β) for both linear and quadratic components and R^2 values, for Category A and B stimuli, respectively. For the Wide condition, the results confirm the significant inverted U pattern for Category A in terms of a negative quadratic coefficient for subjects M and S, as well as for the average data. It is notable that these subjects had higher accuracy levels (*M* = 87%) than the others (*M* = 73%; see Table 7). The quadratic coefficient for subject T was also negative, and approached significance, $p = 0.12$, while the overall regression was significant. This suggests that across subjects, there was a systematic deviation of the obtained Category A data from the predictions of the GLC that could be described approximately as an inverted U pattern. Notably this inverted U pattern is consistent with that obtained with pigeons in Experiment 1 (see Figure 4.6). For Category B, there were no systematic results across subjects. The quadratic component was negative and significant for subject T and the mean data. However, this trend is hard to discern in the scatterplots (Figures 4.12 and 4.13). Results for the Narrow condition confirm the visual inspection of Figures 4.15 and 4.16. There were no significant components in the individual or mean data. Overall, these analyses suggest that results for humans were similar to those obtained with pigeons in Experiment 1: There was an inverted-U shape pattern of accuracy as a function of orientation for Category A in the Wide condition, and results from the Narrow condition showed unidimensional control by frequency and no systematic deviations from the GLC.

EXPERIMENT 1 AND 2 COMBINED RESULTS

Because one of the primary aims of these experiments was to compare the results of humans and pigeons in the same two-dimensional information-integration task, the following section compares the data from the two species directly.

First overall accuracy was averaged across subjects and shown in Figure 4.17. The shaded bars correspond to the human data and the unfilled bars the pigeon data. The wide condition is on the left and the narrow on the right. Bars for each group of subjects and conditions are mean data and standard error bars are included. Overall accuracy was higher for pigeons than for humans in the Wide condition, *M*'s = 89% and 80%, respectively, resulting in a significant *t* statistic, $t(178) = 4.09$, $p<0.05$. However, it is notable that accuracy was higher and comparable to levels obtained with pigeons for the two humans (M and S) whose data showed the significant inverted-U pattern for Category A responding. A likely explanation for the overall increased accuracy of the pigeons was the substantially greater number of trials; pigeons completed an average of 6480 trials compared to 720 trials for the humans. By contrast, in the narrow condition there was no systematic difference in accuracy between species: human subjects were slightly more accurate with 66% correct and the pigeons at 64% correct, despite the difference in amount of training. Additionally this difference did not result in a significant *t* statistic, $t(178) = -0.49$, $p<0.05$.

For a more detailed comparison, we examined performances of humans and pigeons separately for each category. Figure 4.18 shows p(A) for humans plotted against p(A) for pigeons in the Wide condition, separately for Category A (filled squares) and Category B (unfilled triangles). For Category A, the points fall unsystematically around the major diagonal (i.e., line of equality), suggesting that there was no overall difference in performance for pigeons and humans. Consistent with this, the average accuracy for Category A was 81% and 77%, respectively, for pigeons and humans. By contrast, Figure 4.18 shows that humans' performance was overall less accurate than pigeons for Category B. The points fall generally above the major diagonal, and average accuracies for Category B were 94% and 83%, respectively, for pigeons and humans. Thus, accuracies for Category A stimuli were similar, whereas humans were generally less accurate for Category B. Reasons for this difference are unclear.

Similar scatterplots are provided for the Narrow condition in Figure 4.19. For both Category A and B, responding of both species was similar and highly correlated, as evidenced by the points falling unsystematically around the major diagonal in each case. The average accuracies for Category A for pigeons and humans were 61% and 65%, respectively, and for Category B were 71% and 70%.

Figure 4.20 shows average performance for humans and pigeons in the Wide condition, with *p*(A) plotted as a function of orientation for both categories. The inverted-U shaped pattern for Category A responding is apparent for both humans and pigeons. The lower accurate performance of humans for Category B stimuli is evident in that *p*(A) values for humans (unfilled triangles) generally lie above corresponding values for pigeons (crosses).

Overall, these results suggest that performances of humans and pigeons were remarkably similar. The only substantive difference was that humans were overall less accurate in the Wide condition, which could be traced to their performance with Category B stimuli, but not Category A stimuli. Responding to Category A stimuli for both species showed an inverted-U shaped pattern of accuracy as a function of orientation.

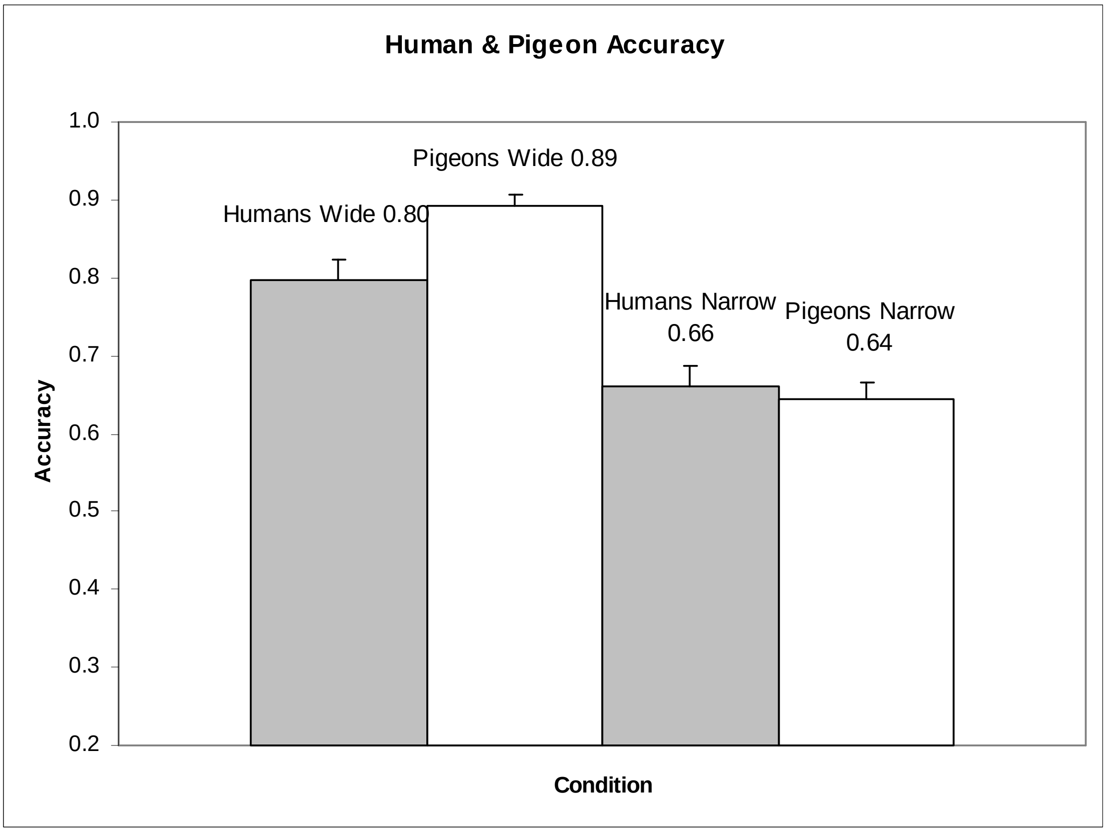


Figure 4.17. Overall accuracy in both the wide and narrow conditions for both the humans and pigeons with standard error bars.

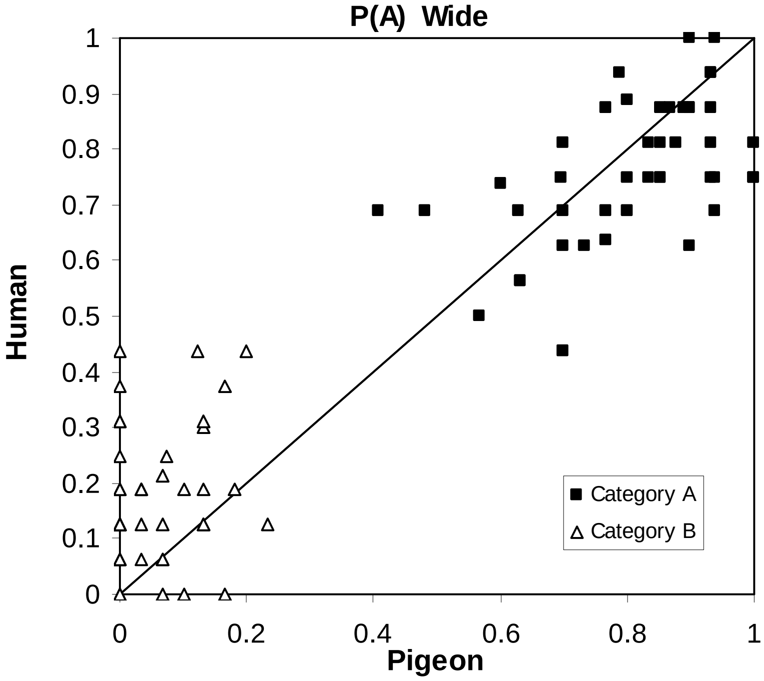


Figure 4.18. Average P(A) in the Wide condition for humans and pigeons plotted against each other on the x and y axis respectively. The optimal decision bound is the solid line along the major diagonal.

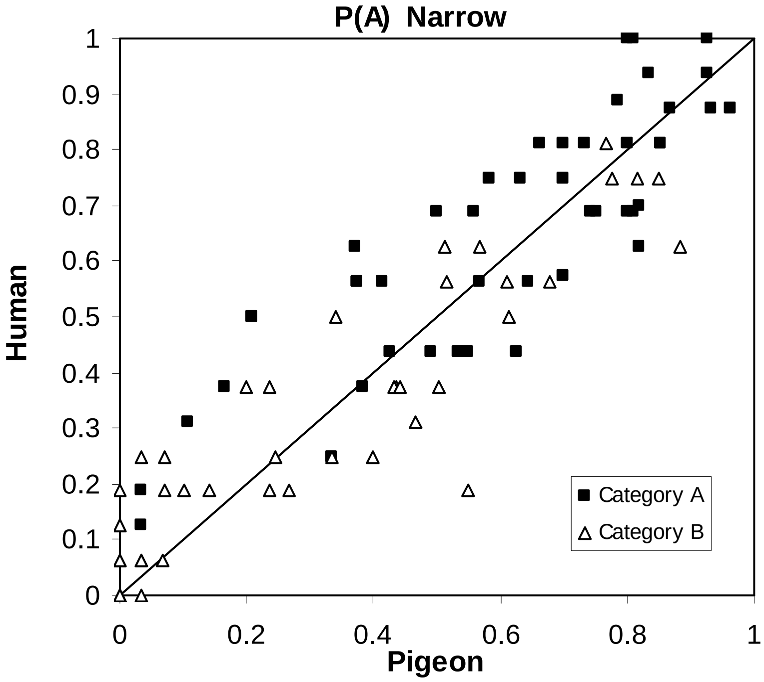


Figure 4.19. Average P(A) in the Narrow condition for humans and pigeons plotted against each other on the x and y axis respectively. The optimal decision bound is the solid line along the major diagonal.

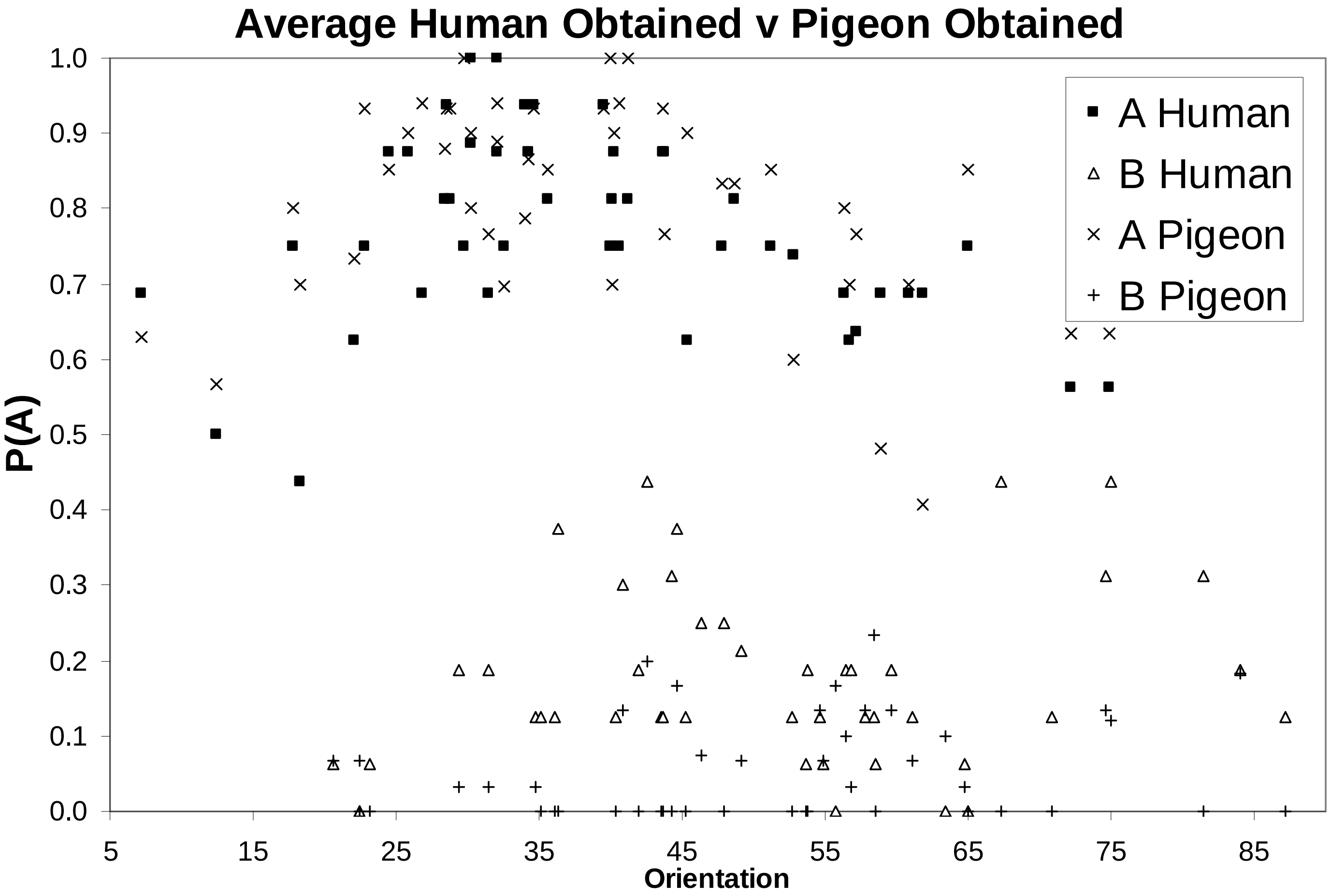


Figure 4.20. Average Human and pigeon probability of category A responding as a function of orientation. Category A human, B human, A pigeon and B pigeon are represented by filled squares, unfilled triangles, x's and crosses respectively.

GENERAL DISCUSSION

The primary goal of these experiments was to compare performance of humans and pigeons on the information integration task developed by Ashby and Gott (1988). In research on human categorization, this task has represented a paradigmatic example of procedural learning, in which subjects must learn categories through trial-and-error rather than the application of explicit rules (Ashby, Alfonso-Reese, Turken & Waldron, 1998). Thus the present research provided an opportunity to test whether humans and pigeons learn visual categories through a similar trial-and-error process. In Experiment 1, pigeons responded in the information-integration task with Gabor stimuli, in which the range of orientation values was either wide or narrow. In Experiment 2, humans responded on the same tasks using exactly the same stimuli. Results showed that performances of pigeons and humans were overall very similar. For both species, accuracy was greater in the Wide than Narrow condition, and model comparison analyses showed that the GLC provided a better fit to the data than unidimensional models for the Wide but not the Narrow condition for all pigeons and humans. In the Narrow condition, the unidimensional frequency model provided a better account of the data, again for all pigeons and humans. This confirms that both species were responding on the basis of both dimensions in the Wide condition, but only frequency in the Narrow condition.

Closer examination of the data showed that the same pattern of systematic deviations from GLC predictions – an inverted-U shaped pattern for Category A responding – was evident for both species. In addition, although accuracy was overall greater for pigeons in the Wide condition, this difference was attributable to Category B responding, where accuracy was lower for humans. There was no systematic difference in accuracy for Category A. Although perfect accuracy was possible, performance of both humans and pigeons was less than optimal in both conditions. Whereas the lower accuracy in the Narrow condition can be attributed to the increased difficulty of the task, the suboptimal performance in the Wide condition was a necessary consequence of the inverted-U shaped pattern of Category A responding. In addition to this pattern, humans also showed overall less accuracy for Category B stimuli, although it is possible that their performance would have increased if they had an equivalent amount of training as the pigeons.

To examine the question of optimality in the information-integration task more generally, we compiled a list of all experiments we could find which have used tasks based on Ashby and Gott's (1988) procedure, in which the stimuli are based on bivariate normal distributions with zero covariances, and have reported accuracy data. In this case, the optimal decision bound is linear, similar to the experiments reported here. These studies are displayed in Table 4.12, which shows for each the number of participants, number of trials, the average percent of correct responses, as well as an effect size measure. The effect size was calculated as the Euclidean distance between the centroids for each category divided by the pooled standard deviation, and was intended as a rough measure of the overall difficulty of the category discrimination in terms of the stimulus parameters. For those experiments that used overlapping category structures in which perfect accuracy was impossible (i.e., some stimuli were closer to the other category centroid), accuracy was defined in terms of the percentage of responses that were consistent with the optimal linear decision bound.

As Table 4.12 shows, optimality was rarely achieved. The only study for which performance was nearly optimal was Ashby and Gott (1988), who used line segments that varied in terms of width and height. As discussed in Chapter 1, these stimuli are problematic because the component dimensions are not fully separable, and both width and height are measured in the same units. Thus subjects may have learned to use a rule such as: Respond Category A if the height is greater than the width, otherwise respond Category B. Use of such a rule may have resulted in higher levels of accuracy. Interestingly, accuracy for Herbranson et al.’s pigeons never reached the same levels as humans in Ashby and Gott’s (1988) study, despite the fact that they used similar stimuli (rectangles) and the pigeons received far more training. The most reasonable conclusion is that pigeons were unable to learn a rule and their asymptotic accuracy was lower. With humans, accuracy levels were lower when stimuli with separable dimensions were used. Ashby and Maddox (1990) conducted an experiment similar to Ashby and Gott (1988) except with semicircles that varied in terms of size and the orientation of a radial spoke as stimuli (i.e., Shepard circles; Shepard, 1964). Overall accuracy levels were much lower in their study – only 81% of responses consistent with the optimal linear decision bound – compared to Ashby and Gott (1988). Because the most salient difference between these experiments was the nature of the stimuli, it suggests that when humans are prevented from using a rule, their performance in the information integration task also falls short of optimality.

To investigate whether our effect size measure might predict performance in the II task, we examined whether accuracy levels were correlated with effect size for studies that used non-overlapping categories. Studies with overlapping categories were excluded because some used non-separable stimuli (Ashby & Gott, 1988). Across studies, there was a small positive correlation, r = .18, which failed to reach significance. Although this provides no evidence that performance varied with the difficulty of the task (as measured by the effect size), given the relatively small number of studies, and the other potential confounds (e.g., pigeons vs humans), the question of how the difficulty of the II task may vary depending on category distributions requires further study.

Source	Experiment	# Participants	# Trials	% Correct	Stimuli	Effect Size
Ashby & Gott 1988	Exp1	3	300	96*	Line length & width	1.68
Ashby & Gott 1988	Exp 3	3	300	91*	Line length & width	0.87
Ashby & Maddox 1990	Exp 3	5	300	81*	Semi circles and radial line	1.68
Herbranson, Fremouw & Shimp 1999	Exp1/1	4	5400	88*	Rectangles width & height	3.29
Maddox, Ashby & Gottlob 1998	LI	42	100	87	Circle diameter and radial line	3.35
Ashby, Queller & Berretty 1999	Exp 1a	5	800	67	Line length & orientation	2.39
Maddox, Ashby & Bohil 2003	Exp 1	10	320	68	Gabor patch O & F	1.17
Our CH 1	II Condition	6	1260	83	Gabor patch O & F	1.33
Our CH 2	Exp 1 Pigeon W	3	6480	89	Gabor patch O & F	1.52
Our CH 2	Exp 2 Human W	4	720	80	Gabor patch O & F	1.52

Table 4.12. For each of the sources displayed on the left, Experiment, number of participants and trials, percent correct, stimuli and calculated effect size is displayed. Percent correct values marked with an asterisk indicate that category distributions overlapped making perfect performance impossible. In these cases the value indicates the percent of responses consistent with the optimal linear decision bound.

Comparison of the Narrow and Wide conditions showed that control by orientation (and accuracy) decreased when the range of orientation values was reduced from 80° to 10°. This contrasts with research with unidimensional stimuli which has shown that restrictions in stimulus range produce an increase in sensitivity in successive discriminations (Hinson & Lockhead, 1986). In the psychophysics literature, these have been known as *range effects*, and have sometimes been interpreted in terms of limitations in working memory capacity (Gravetter & Lockhead, 1973; Miller, 1956). By contrast, in the Narrow condition when the range of orientation values was restricted there was a reduction in control by that dimension. This suggests that different dimensions may compete for attentional resources in the II task. If differences along one dimension are more salient than the other, it may acquire exclusive control over responding. This may have occurred for frequency in the Narrow condition.

An interesting result was that although pigeons performed overall better in the Wide condition than humans, accuracy levels for pigeons and humans were comparable in the Narrow condition. This shows that humans' learning was more efficient in the Narrow condition because they received far fewer training trials than the pigeons did. Because model comparison analyses showed that responding for both species was controlled solely by frequency in the Narrow condition, a possible reason for the humans' greater efficiency was that they were using a rule-based strategy. This result would be consistent with Smith, Minda and Washburn (2004), who found that humans could learn categories that could be described in terms of verbal rules faster.

An important difference between the present research and previous studies is that we examined asymptotic performance in detail, specifically in terms of whether accuracy varied systematically as a function of stimulus characteristics. We plotted accuracy as a function of orientation, although we could have just as well examined performance as a function of frequency because the dimensions were strongly correlated.

Overall, the present results provide additional evidence that performance of humans and nonhumans is comparable when learning categories that cannot be described in terms of verbal rules. The similarity of the data for pigeons and humans suggests an intriguing possibility: Organisms may share a system for learning visual categories on the basis of experience that evolved millions of years ago, prior to the divergence of avians and mammals. However, the failure of the GLC to account for the data suggests that our results may pose a challenge for models of category learning. Whether other theoretical accounts, such as exemplar and prototype theory, can predict these results is considered in the next chapter.

Chapter 4 - General Discussion: The Fuzzy Prototype (FP) Model

The purpose of this chapter is to consider the theoretical implications of the empirical work in Chapters 2 and 3 for our understanding of category learning and serve as a General Discussion for the thesis. Because one of the major questions at the outset was whether organisms are capable of information integration in tasks using stimuli with fully-separable dimensions, we begin by reviewing Massaro and Friedman's (1990) definition of information integration. We then summarize the key empirical results from previous experiments in this thesis, and ask whether they are consistent with current models for category learning. We will consider the General Linear Classifier and decision-bound models (Ashby, 1992; Ashby & Alfonso-Reese, 1995); prototype models (Posner & Keele, 1968); exemplar models (Nosofsky, 1984, 1986); and the striatal pattern classifier (Ashby & Waldron, 1999). As we will show, our results are not predicted by current models, and so we propose a new model, which we call the 'fuzzy prototype' model. We show that this model is able to account for the major patterns in the data reported in this thesis, and holds promise as an account of category learning in both humans and non-humans.

Information integration

Massaro and Friedman (1990) provided a definition of information integration in the context of a General Stage Theory for conditional discrimination. According to their view, accurate performance requires three stages: Evaluation, Integration, and Decision. In the Evaluation stage, physical stimuli are translated into their equivalent psychological representations. During the Integration stage, some combination of representations is effected. Finally, in the Evaluation stage some decision rule is implemented and the subject makes a response.

Based on this general framework, Massaro and Friedman (1990) defined *integration models* as those which assumed that some combination of representations from different dimensions occurred during the Integration stage. By contrast, in a *non-integration model*, the output of the Evaluation stage is assumed to feed directly to the Decision stage.

Summary of key empirical results

The ‘information integration’ task developed by Ashby and Gott (1988) and used extensively in subsequent research by Ashby, Maddox and colleagues (e.g., Maddox & Ashby, 2004) was employed in experiments reported in Chapters 2 and 3. In Chapter 2, pigeons responded in an information integration task with stimuli modelled closely on those used by Maddox et al. (2003); in Chapter 3, both pigeons and humans were exposed to an information integration task in which the range of orientation values was either wide or narrow across different conditions. Stimuli used in the task for Experiment 2 (wide condition) are shown in Figure 4.1. The task is described as ‘information integration’ because correct responding requires attention to both orientation and frequency. The key result, obtained in Chapter 2 (and replicated within the experiment) for pigeons, and in Chapter 3 for both pigeons and humans, is that the probability of a correct response for Category A (designated as the stimuli with overall higher orientation values in Figure 4.1) was an inverted-U shaped function of orientation, whereas accuracy of responding for Category B did not vary systematically with orientation.

For the RB task, there was evidence of an interaction between orientation and frequency in Chapter 2. Specifically, accuracy was higher in the frequency-relevant RB task when the orientation was relatively low (i.e., near horizontal lines), compared to when it was relatively high (i.e., near vertical).

Assessment of Current Models for Category Learning

As we now show, results from both the information-integration and rule-based tasks pose a challenge to current models for category learning. We first consider the General Linear Classifier (GLC). As noted in Chapters 2 and 3, the systematic pattern in the residuals confirms that the GLC is unable to account for the inverted-U shape in the information-integration task. However, these results are problematic for any decision-bound model. The basic assumption of this class of models is that organisms learn a ‘boundary’ in the internally-represented stimulus space, such that stimuli are classified as one category or the other depending on which side of the boundary that a stimulus presented on a trial lies. The probability of responding to a category depends on its distance from the boundary. Particularly troublesome for decision-bound models is the asymmetry of our results: The inverted-U shape was obtained for Category A, but not Category B. Note that by itself, the inverted-U shape could be accommodated by a decision-bound model – for example, the General Quadratic Classifier, which allows for a non-linear (i.e., curved) decision boundary. To account for the inverted-U pattern, the boundary would need to be shaped like a crescent such that mid-range Category A stimuli lie above the boundary, whereas stimuli on either end were closer to the boundary. However, the problem is that accuracy for Category B would then need to be higher for stimuli with relatively low or high orientation values. In general, any decision bound will require that accuracy vary symmetrically, because as the boundary moves closer to stimuli in one category, it necessarily moves away from the other. Thus, the asymmetry in our results poses a problem for decision-bound models in general.

According to prototype models (e.g., Posner & Keele, 1978), subjects learn to associate a most typical or representative exemplar (‘prototype’) with each category response. When a stimulus is presented on a trial, subjects compute the similarity of the stimulus to each prototype, and the probability of responding is then determined by the ratio of the similarities. Prototype models are unable to account for several features of our data. First, if the prototype is assumed to be the centroid of each category, prototype models would predict that accuracy should decrease within each category as the orientation value moves away from the centroid. That is, there should be a peak in accuracy near the centroid (prototype), and a symmetrical decline on either side. However, the inverted-U shape pattern that we observed had a plateau, not a well-defined peak. The prototype model is also unable to account for the relatively constant accuracy for Category B. Thus the prototype model fails to provide a full explanation of our results.

In exemplar models, subjects store a number of representative stimuli (‘exemplars’) from each category in memory (Brooks, 1978; Nosofsky, 1986). When a stimulus is presented on a trial, the similarity between the stimulus and each exemplar is computed for both categories. The similarity scores are averaged within categories, and the probability of responding is determined by the relative average similarity. Ashby and Alfonso-Reese (1995) showed that exemplar models were similar to decision-bound models such as the GLC, in that both models were equivalent to a process in which the observer estimated the likelihood that the stimulus presented on a trial was sampled from each category distribution. However, Ashby and Alfonso-Reese (1995) distinguished between parametric and non-parametric classifiers. A parametric classifier assumes that subjects must estimate parameters (i.e., means, variances, covariances) of the stimulus distributions associated with each category. For example, the GLC is a parametric classifier, because the subject is taken to assume that the stimulus distributions are multivariate normal, with zero covariance. Under these conditions, the optimal decision bound is linear (Ashby & Gott, 1988). Thus because the subject must estimate the parameters of the distributions in order to compute the optimal decision bound, the GLC is a parametric classifier. By contrast, Ashby and Alfonso-Reese (1995) noted that exemplar models were non-parametric classifiers, because subjects were not assumed to estimate distributional parameters. However, Ashby and Alfonso-Reese showed that exemplar models were equivalent to a process in which subjects estimated the likelihood that a stimulus was drawn from the distributions for each category essentially by using a relative-frequency histogram (specifically, a Parzen kernel estimator).

Applied to the information integration task in Figure 4, exemplar models make equivalent predictions as the GLC because they assume that subjects should learn the stimuli in each category, and eventually reach optimal performance. Exemplar models have no *a priori* basis to predict that suboptimal performance should vary systematically with orientation; specifically that the inverted-U shape pattern should be obtained for Category A, but not the corresponding pattern for Category B. Thus our results also pose a challenge for exemplar models.

The final model to consider is the striatal pattern classifier (SPC; Ashby & Waldron, 1999), which has been used as the ‘procedural learning’ component in multiple systems neurobiological theories proposed by Ashby and colleagues, including COVIS (Ashby, Alfonso-Reese, Turken, & Waldron, 1998) and SPEED (Ashby, Ennis, & Spiering, 2007). According to the SPC, stimuli are represented in a perceptual space, which corresponds to a region in the striatum. Because there are many projections from visual cortex that converge on the striatum (Saint-Cyr, Ungerleider, & Desimone, 1990), Ashby and Waldron (1999) proposed that the striatum provides a low-resolution representation of the perceptual space. The model assumes that different cells (and hence regions of the perceptual space) become associated with different category responses through reinforcement learning. Although the SPC provides a neurobiologically-plausible account of category learning, similar to exemplar theory it provides no rationale for predicting the inverted U-shaped pattern that we observed. The model cannot explain why the accuracy should decline at either extreme of the orientation dimension, and why a corresponding change in accuracy does not also occur for Category B.

The other result that is problematic for current models of category learning was obtained in the rule-based (RB) task. In the RB condition in Experiment 1 and 2 of Chapter 2, an interaction was obtained between frequency and orientation, such that sensitivity to frequency (and accuracy) was greater when the orientation value was relatively low (i.e., lines close to horizontal) than when it was relatively high (i.e., lines close to vertical). Apart from making post-hoc assumptions, none of the models for category learning described above – the GLC, prototype model, exemplar model, and SPC – can predict this result. Next we consider whether other models might be able to provide a more satisfactory account of our data.

Associative Model

First we pursued alternative models which we hoped might yield a ‘non-integration’ model, in Massaro and Friedman’s (1990) terms, which could successfully account for performance in the information-integration task. The most promising approach was to assume that subjects learned to associate extreme values on a particular category with a specific response. The key feature of the model was that critical values on different dimensions were associated with the different category responses. For example, a relatively high frequency value might be associated with the Category A response, whereas a relatively high orientation value might be associated with the Category B response. We then assumed that the similarity of the percept’s orientation and frequency to these extreme dimensions would determine the probability of a Category A response. In Massaro and Friedman's terms, this would be a non-integration model because the different response tendencies were competing with each other. When the frequency was high and the orientation was low (i.e., the upper left corner of Figure 5.5), the probability of a Category A response would be maximal; whereas when the frequency was low and the orientation was high (i.e., lower right corner of Figure 5.5), the probability of a Category A response would be minimal.

As it turned out, this model yielded a description of the data that for all practical purposes was equivalent to the GLC. To see how this occurred, imagine points on the major diagonal in Figure 5 (i.e., decision line for GLC). For these points, the orientation and frequency are equally close to their extreme values (i.e., values associated with the prototypes), and so the subject should be indifferent between category A and B. However, if the point moves either horizontally or vertically – so that it becomes closer to one extreme value but not the other – then the subject will be more likely to choose one of the categories. For example, if the point moves towards the right, the orientation becomes closer to the extreme value associated with Category B, and hence choice for Category A decreases.

At first we thought that this model might provide a more parsimonious account of performance than the GLC because it was a non-integration model. However, because its predictions were virtually identical to those of the GLC when fitted to the data, it had exactly the same failure in terms of being unable to account for the inverted-U shaped pattern in the Category A responding, and the interaction in the RB condition. Also like the GLC, it predicted that accuracy for the two categories would show corresponding changes, and thus could not account for the asymmetry in the results (i.e., the lack of systematic change in accuracy for Category B).

Summary

Given that existing models for category learning, as well as the associative model, are unable to account for these findings, the question becomes whether an alternative model might be more successful. In particular, from a comparative perspective the present results suggest that perceptual category learning in humans and pigeons may be based on a similar process, at least for the information integration task. The question we address now is how that process might be characterized.

Fuzzy Prototype Model

Similar to previous models, we assume that stimulus percepts in the Gabor task can be represented in a two-dimensional stimulus space, such as that in Figure 5.5. This figure shows the stimulus categories from the information integration task used in earlier chapters in this thesis. The basic idea of the new model is very simple: Through a process of trial-and-error reinforcement learning, a contiguous region of the perceptual space becomes gradually associated with one of the categories. Note that this assumption is similar to the SPC (Ashby & Waldron, 1999); below we consider differences between the FP model and other models for category learning. This region which we will represent for sake of simplicity as a line segment in the stimulus space is referred to as a ‘fuzzy prototype’. Importantly, only one such prototype is assumed to be learned, and is associated with only one of the two categories. In what follows, we will assume that the prototype is associated with Category A, although there is no reason that it could not be Category B instead. The important point is that the subject learns to associate one prototype with a particular response, and that prototype is a linear segment (i.e., fuzzy) of the stimulus space. When a stimulus is presented on a trial, the probability that the subject responds ‘Category A’ is determined by the similarity of the percept with the fuzzy prototype. Similarity is calculated as a function of the minimum distance between the percept and the fuzzy prototype. Specifically, the model computes the minimum Euclidean distance from the point which represents the percept to the line segment. The probability of responding ‘Category A’ is then a decreasing function of the distance.

Several additional assumptions are required for the model to make quantitative predictions. First, because orientation and frequency are measured in different physical units, values for both dimensions were standardized (i.e., *Z* scores). Note that an alternative might be to use scale factors for each dimension, but because of the added complexity we did not pursue this approach here. Second, the coordinates for the end points of the line segment defining the fuzzy prototype were assumed to be parameters in the model; once these were defined, five equally-spaced interior points were computed. Thus the fuzzy prototype was represented by 7 co-linear points in the stimulus space. Note that the number of interior points is arbitrary, and the fuzzy prototype could be defined analytically in terms of a linear equation. However, we believe that a finite number of points provides a more realistic model of how the prototype might be represented in visual cortex. The minimum distance is then computed as the minimum of the distances between the stimulus presented on a given trial, and the 7 points comprising the prototype. Finally, the probability of responding ‘Category A’ is calculated as a decreasing power function of the minimum distance.

Specifically, the odds ratio for responding Category A is a negative power function of the minimum distance:

$$\frac{p(A)}{p(\sim A)} = D^{-q} \text{, which implies that } p(A) = \frac{D^{-q}}{D^{-q}+1}$$

(1)

Application of the Fuzzy Prototype Model to the Data

Next we apply the fuzzy prototype (FP) model to the data from the empirical chapters of this thesis. The major question is whether the FP model can provide an adequate account of the data, unlike current models for category learning. The FP model contains five fitted parameters: The orientation and frequency coordinates (standardized) for the two endpoints of the line segment, and the power function exponent, q. The FP model was applied to group-mean data from the following conditions: Chapter 2: Experiment 1: II condition, RB condition (frequency-relevant), II replication condition; Experiment 2: RB frequency-relevant; RB orientation-relevant. Chapter 3: Humans, II-wide condition; Pigeons, II-wide condition. In all cases, parameters were obtained using maximum likelihood estimation. A nonlinear optimization program (Microsoft Excel Solver) was used to obtain the parameter estimates.

Chapter 3 – Experiment 1

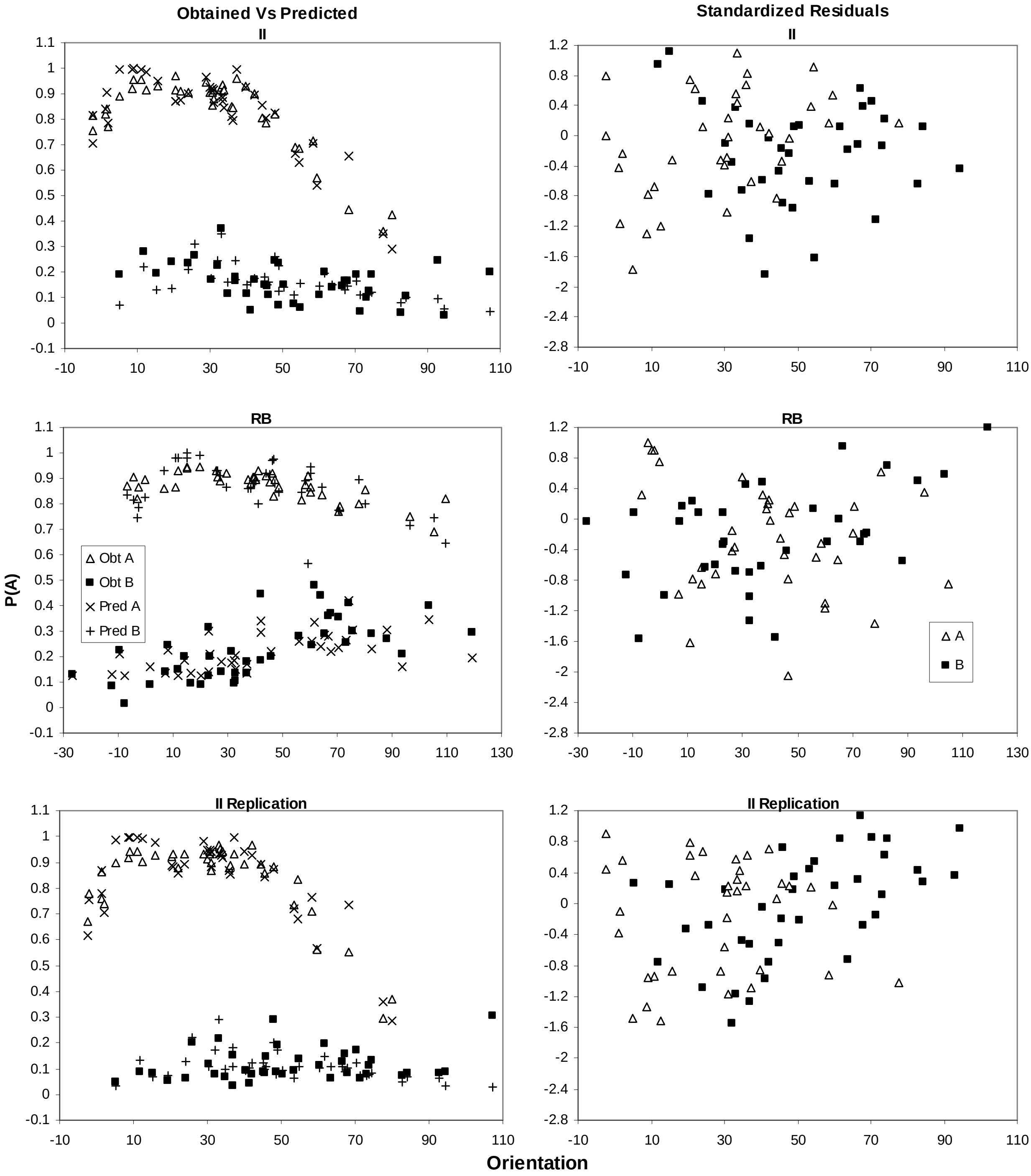


Figure 5.1. Obtained probability of responding category a (P(A) plotted against orientation. The left column shows the obtained data in the unfilled triangles and filled squares for category a and b respectively. Also plotted are the values predicted by the FP with crosses and x symbols. The right hand column shows the standardized residuals from fitting the FP.

Chapter 3
Wide

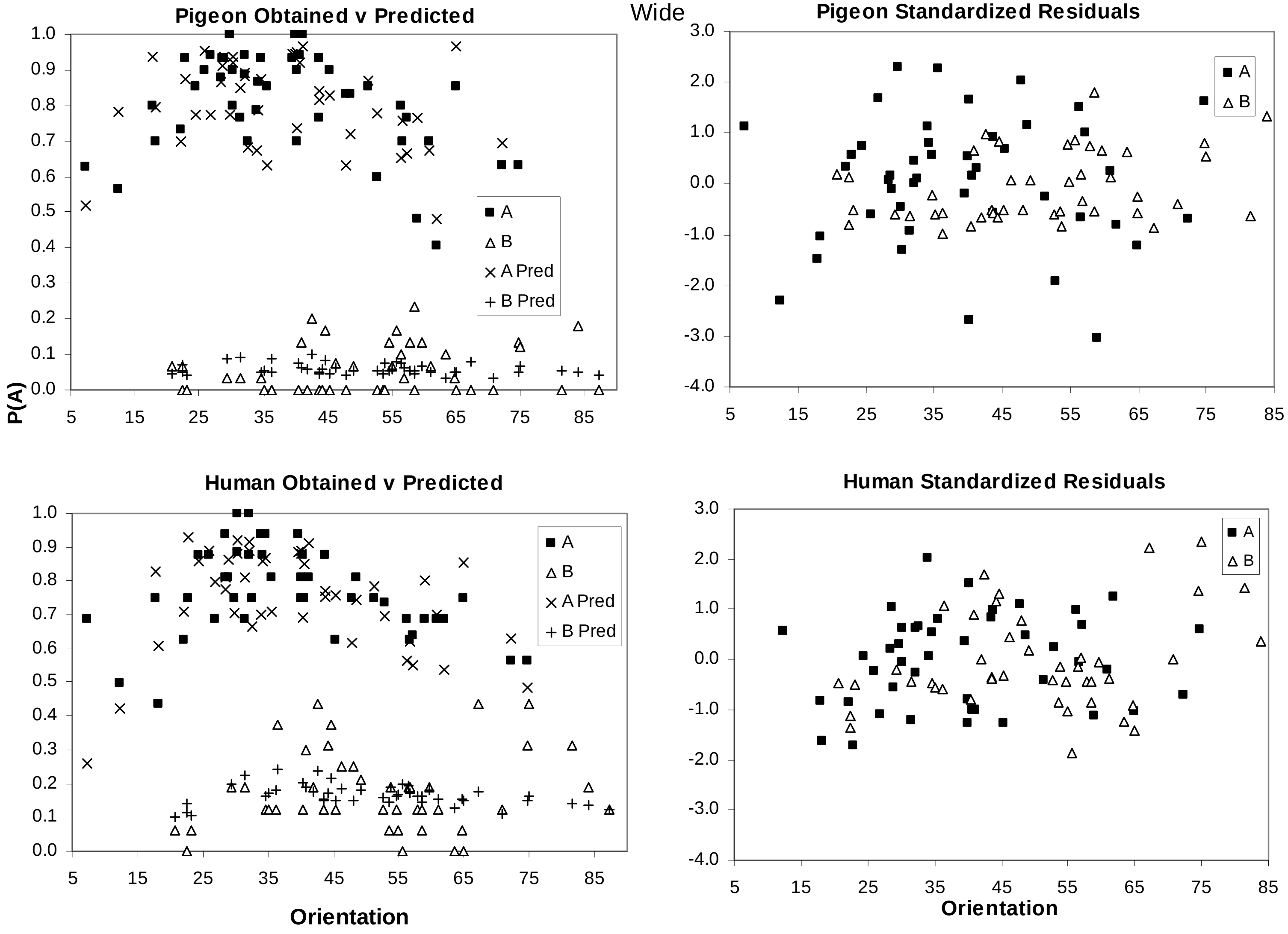


Figure 5.2. The left panels show the average obtained vs. predicted values for Pigeons (top) and Humans (bottom) and the right panels display the average standardized residuals from the FP model fits.

Chapter 2 - Experiment 2

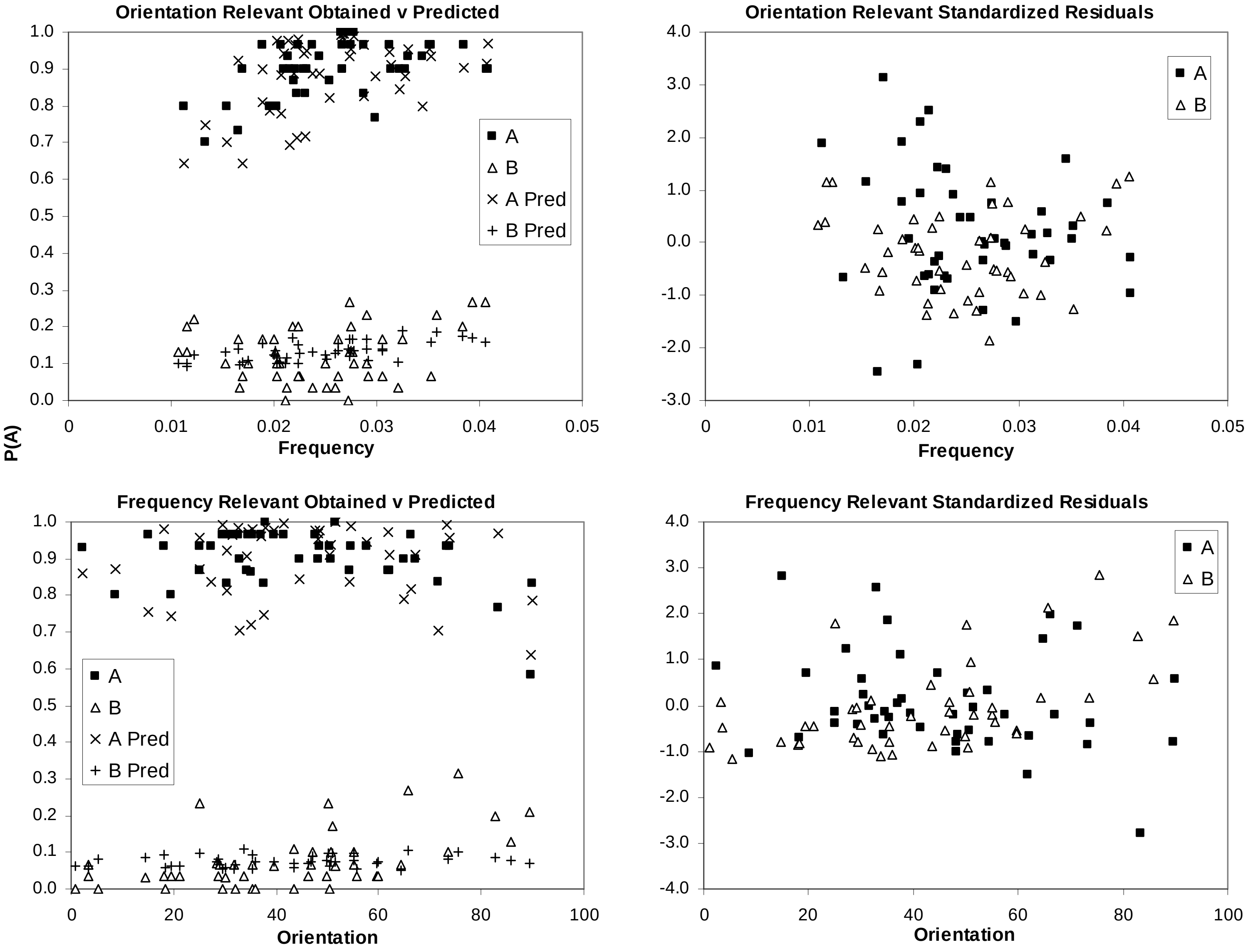


Figure 5.3. The left column shows the obtained probability of responding category a (P(A) plotted against orientation for the Orientation (top) and Frequency (bottom) conditions. Obtained data is shown in the unfilled triangles and filled squares for category a and b respectively. Also plotted are the values predicted by the FP with crosses and x symbols. The right hand column shows the standardized residuals from fitting the FP.

		Ch1 Exp 1			Ch2		Ch1 Exp 2	
		II	RB	II Rep	Pigeon	Human	RB F	RB O
FP	VAC	0.98	0.95	0.98	0.94	0.88	0.97	0.96
	AIC	77.13	83.46	69.39	68.84	91.87	56.58	70.17
GLC	VAC	0.89	0.85	0.88	0.93	0.70	0.94	0.93
	AIC	67.66	75.15	66.96	63.71	84.86	53.56	67.58

Table 5.1. VAC and AIC values for the FP and GLC models, fitted to data from each experiment.

	Category A			**Category B**		
FP	β Linear	β Quadratic	R^2	β Linear	β Quadratic	R^2
II CH1 Exp 1	-0.23	0.70 ***	0.49 ***	0.03	-0.15	0.02
RB CH1 Exp 1	0.12	-0.21	0.05	-0.21	0.2	0.06
II Rep CH1 Exp 1	-0.05	-0.26	0.08	0.38 **	0.40 **	0.34 ***
Human CH2	-0.1	0.15	0.24	0.26	0.05	0.08
Pigeon CH2	0.02	-0.15	0.02	0.28	-0.06	0.08
RB O CH1 Exp 2	0.03	-0.15	0.02	0.28	-0.06	0.08
RB F CH1 Exp 2	-0.24	-0.01	0.06	0.53 ***	0.25 *	0.38 ***

Table 5.2. Regression results for the FP model in each of the experiments. Asterisks indicate significance levels. Significant regression results for β Linear, β Quadratic and R^2 values are indicated with an asterisk (*) for 0.1, 0.05, 0.01 significance levels using *, ** and *** respectively.

FP	Orient 1	Freq 1	Orient 7	Freq 7	QExp
II CH1 Exp 1	2.26	0.02	48.11	0.04	3.08
RB CH1 Exp 1	-54.51	0.03	110.45	0.03	1.99
II Rep CH1 Exp 1	1.24	0.02	49.15	0.04	3.65
Human CH2	17.87	0.02	67.09	0.04	2.82
Pigeon CH2	9.33	0.02	82.51	0.05	4.63
RB O CH1 Exp 2	29.87	0.02	34.51	0.05	2.33
RB F CH1 Exp 2	29.87	0.02	34.51	0.05	2.33

Table 5.3. Model fit parameters, Orient 1 and Freq 1 are coordinates of one endpoint of the model; whereas orient7 and freq7 are the coordinates for the other. QExp is the exponent in the power function in Equation 1.

Scatterplots of obtained versus predicted data, and standardized residuals, are shown for FP model fits to data from Chapter 2 – Experiment 1, Chapter 3 and Chapter 2 – Experiment 2, respectively, in Figures 1 through 3. Parameter estimates and variance accounted for, and likelihood statistics are provided in Table 5.1.

First we will consider the FP fits to the information integration condition data, specifically Chapters 2 – II condition and replication, and Chapter 3, humans and pigeons. As Figures 5.1 and 5.2 show, in each of these data sets, the inverted-U shaped pattern for response accuracy for Category A stimuli was accurately predicted by the FP model. Also, the model predicted that accuracy for Category B stimuli was approximately constant across the range of orientation values. The ability of the FP model to capture the patterns in the data is confirmed by the goodness-of-fit statistics. The percentage of variance accounted for by the FP model was higher than that achieved by the GLC in each case: In both conditions of Chapter 2, the FP model accounted for 98% of the variance, and in Chapter 3 accounted for 94% and 88%, respectively, of the pigeon and human data. These results provide evidence that the FP model is able to account for the major features of performance in the information-integration conditions.

To provide a more detailed examination of the FP model fits, we performed a series of polynomial regressions in which the standardized residuals were regressed on the orientation values. As in the previous chapters, the orientation values were centred prior to the regressions. Those regressions showed that the residuals from the GLC predictions deviated systematically as a function of orientation: For the average data, the Category A residuals showed a significant negative quadratic component (resulting from the GLC's inability to predict the inverted-U shaped pattern) and a significant positive linear component (resulting from the GLC's inability to predict that accuracy was approximately constant for Category B). These results were obtained for both the II and replication conditions of Chapter 2, and for both humans and pigeons in the wide condition in Chapter 3.

By contrast, regressions of the standardized residuals from FP model fits on orientation values revealed no systematic deviation across conditions. These results are shown in Table 5.2. There was a significant positive quadratic component for Category A in the II condition of Chapter 2, but not in the replication condition and in neither of the conditions from Chapter 3. For Category B, the linear and quadratic components were significant in the II replication condition (Chapter 2), but not in the II condition or in the Chapter 3 conditions. Because of the likelihood of Type I error due to the number of statistical tests, and the failure to replicate any significant component – in contrast to the GLC results – we conclude that overall, the predictions of the FP model did not deviate systematically from the predicted values. This result, in combination with the excellent goodness-of-fit statistics and its ability to capture the qualitative patterns in the data, suggests that the FP model provides a superior account of responding in the information-integration task compared with other current models.

Next we consider the fits of the FP model to the data from the RB conditions. As shown in Chapter 2, accuracy in the RB – frequency relevant task was greater when the orientation was relatively low (i.e., close to horizontal) than when it was relatively high (i.e., close to vertical), suggesting an interaction between frequency and orientation. As Figure 5.1 shows, the FP model was able to predict this interaction, in both experiments of Chapter 2. The variance accounted for by the FP model was 95% and 97%, respectively, in these conditions, confirming that the model was able to capture the major features of the data. Polynomial regressions on the standardized residuals found that no components were significant in the RB – frequency relevant condition from Experiment 1 for either category. The linear and quadratic components were positive and significant for Category B for the RB – frequency-relevant condition in Experiment 2. This pattern can be observed in the lower right panel of Figure 5.3: For Category B stimuli, the standardized residuals showed a U-shaped pattern (positive quadratic). Although reasons why this pattern was obtained are unclear, it was not observed in Experiment 1 and thus may have been due to chance.

By contrast, the GLC was unable to account for the interaction in the RB – frequency-relevant data. For both RB frequency-relevant conditions, the regression analyses revealed significant positive linear components for Category A, and negative linear components for Category B (see Figures 2.7 – Experiment 1, and Figure 3.5 – Experiment 2, in Chapter 2). Thus, the deviations in the GLC predictions were systematic, whereas those for the FP model were not.

Somewhat surprisingly, no similar interaction was obtained in the RB – orientation relevant condition. Accuracy was approximately constant for both categories as a function of frequency (see Figure 5.3). The FP model accounted for a high proportion of the variance (96%), and regressions on the standardized residuals found no systematic deviations. Although reasons why the interaction between frequency and orientation was obtained in the frequency-relevant but not orientation-relevant condition are unclear, the FP model provided a good account of performance in both RB conditions.

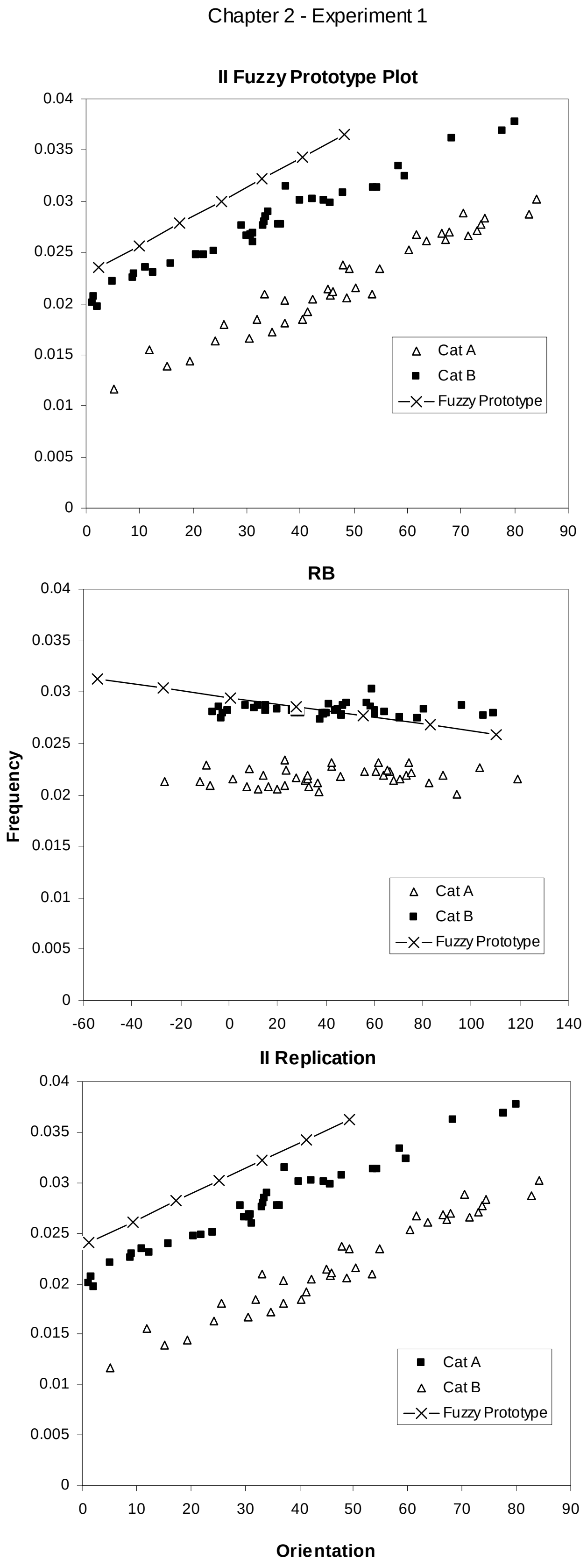


Figure 5.4. Stimuli plots (orientation x, frequency y axis) for the II, RB and II replication conditions, FP model predictions are displayed by connected ‘x’ symbols.

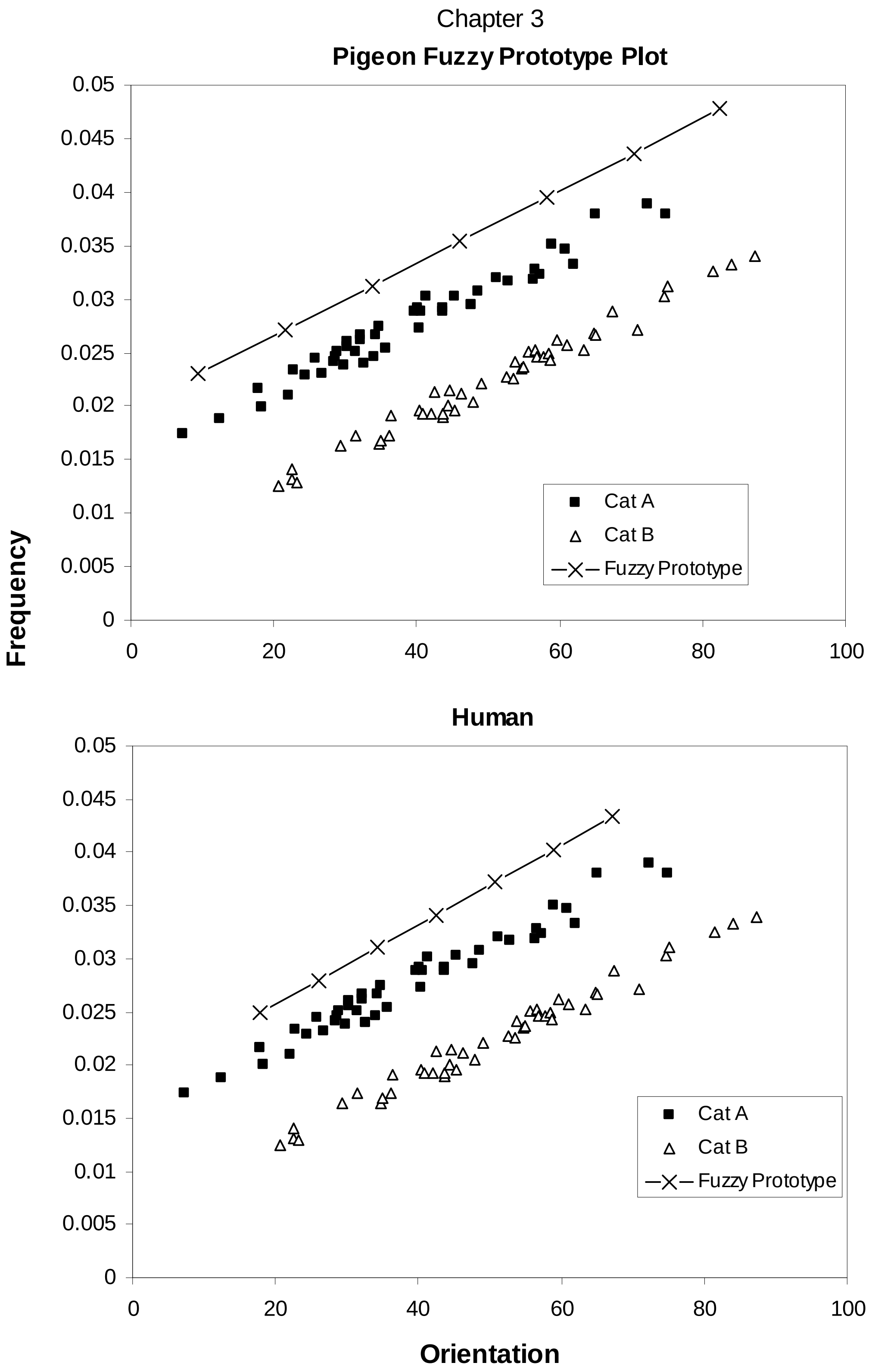


Figure 5.5. Stimuli plots (orientation x, frequency y axis) for pigeon and human conditions, FP model predictions are displayed by connected ‘x’ symbols.

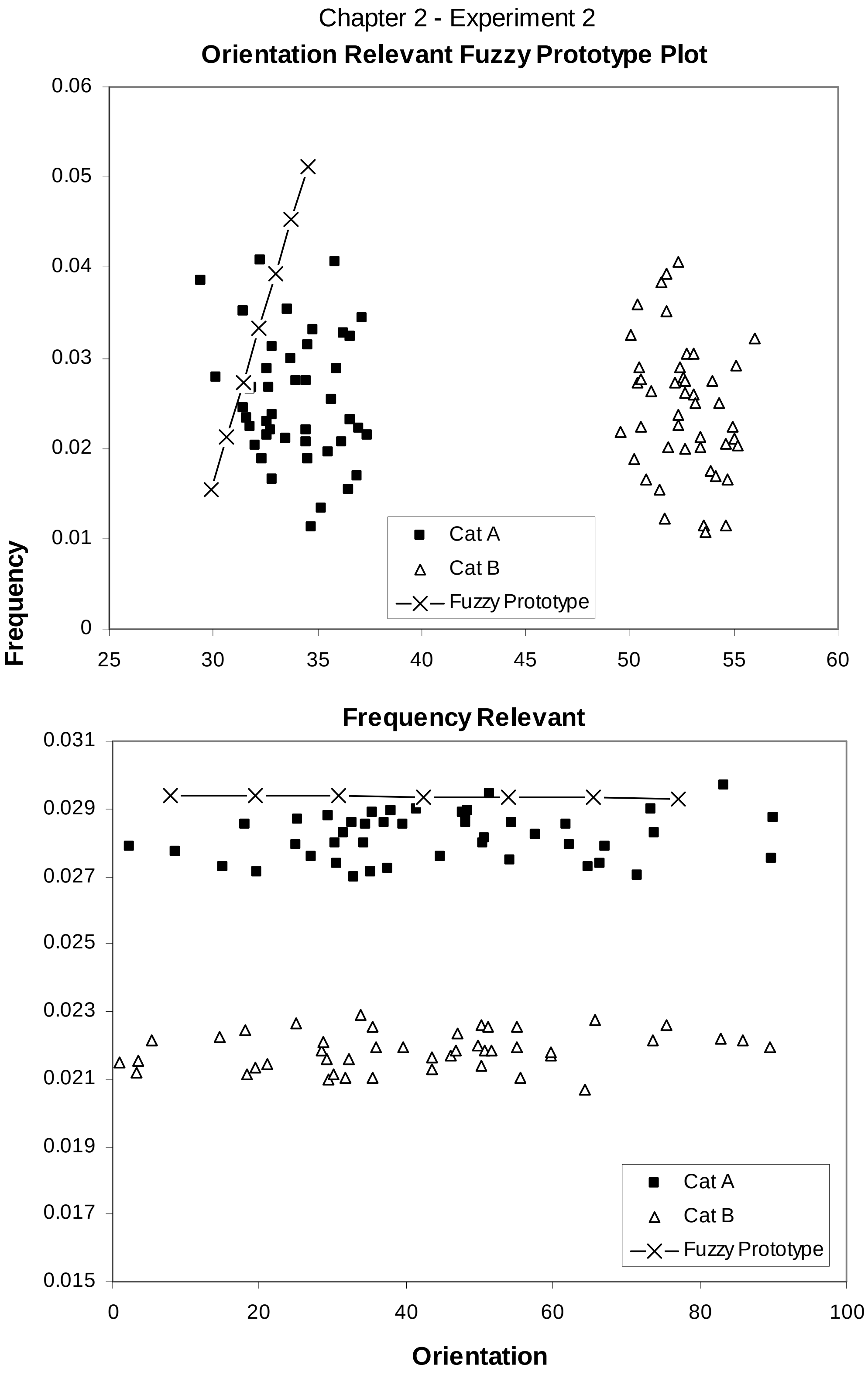


Figure 5.6. Stimuli plots (orientation x, frequency y axis) for the orientation and frequency relevant conditions, FP model predictions are displayed by connected ‘x’ symbols.

Some insight into how the FP model was able to account for the data from the present experiments can be gained from Figures 5.4 through 5.6, which show the category stimuli from each condition plotted in a two-dimensional stimulus space. For each condition, the fuzzy prototype associated with the model fit – the line segment in the stimulus space – is also shown. First we consider performance in the information-integration tasks.

By convention, the category with the overall higher frequency values was designated as Category A in the information integration conditions. In each case, the fuzzy prototype was associated with (i.e., closer to) Category A. The reason for this, in terms of the model fits, is straightforward: The FP model will predict the inverted-U shape for the category associated with the prototype, to the extent that the prototype does not extend over the full range of the category stimuli. Because the inverted-U shape was always obtained for pigeons and humans for Category A but not Category B, the fuzzy prototype was always associated with Category A. However, just why this occurred is unclear. In terms of the model, there is no *a priori* reason to expect that the fuzzy prototype would not be associated with Category B. Thus, an important question for future research will be to explore conditions in which prototypes might be learned to specific categories. For example, subjects could be given pre-training with one of the categories but not the other, and later exposed to the full information-integration task. Presumably, subjects pre-trained with Category B rather than Category A would be predisposed to associate the fuzzy prototype with Category B (and vice versa).

Another salient aspect of the FP fits to the information-integration conditions is that the fuzzy prototype did not superimpose on the Category A stimuli, but rather was somewhat above and to the left (see Figures 5.4 and 5.5). Thus, none of the actual stimuli were equivalent to the prototype. Whether this is a realistic assumption – in the sense that the estimated prototype corresponds to subjects' learning, or in more cognitivist terms, to how the prototype is represented in the brain – is an open question. However, the prototype's displacement ensured that there was at least some distance between the Category A stimuli and the prototype. If the prototype had overlapped the stimuli, then the distances for some of the stimuli would have been close to zero, implying near-perfect accuracy. Yet results showed that accuracy, though high, fell short of optimality: The average accuracy for Category A stimuli in the peak performance range (i.e., the top of the inverted-U shape) was approximately 90%.

A future study could use a transfer design in which stimuli that are calculated to lie on the fuzzy prototype, based on fits of the model to baseline data, could be presented in test trials. If the fuzzy prototype is an accurate representation of learning in the task, then performance on these test trials should be superior to that on baseline trials (assuming that comparable levels of suboptimal accuracy were obtained as here).

If the displacement of the fuzzy prototype away from the Category A stimuli is due to the need for the model to accommodate non-optimal performance, another way to achieve this might be to include terms to represent different sources of variance that might affect the decision process, such as perceptual or criterial variance. In functional terms, including one of more of these sources of variance means that it is never (or rarely) possible to have a stimulus that has a distance of zero from the fuzzy prototype.

Perceptual and criterial variance have been included as part of General Recognition Theory (GRT), which is the family of models that includes the GLC (Ashby & Townshend, 1986). However, Ashby has noted that it is very difficult to distinguish between perceptual and criterial variance empirically, and in most applications, they have been subsumed into a single parameter. A variance term corresponds to the assumption that the stimulus, as represented and processed by the organism, likely differs from that presented by the experimenter; the variance quantifies the degree of uncertainty and is usually modelled as a Gaussian random variate. With respect to the fuzzy prototype model, there are three likely sources of variance: Variance might affect the prototype (so that the effective prototype on any given trial was not constant, but could be characterized as a random variable with a specified distribution), the stimulus percept, or the perceived distance between the stimulus and the prototype. Variance in the distance would presumably reflect the sum of prototype and stimulus variance, as well as any variance unique to the computation of distance, and so would likely be the largest source of variance in the system. Including these sources of variance would make the model more complicated, but would be necessary in case future research showed that performance decreased as transfer stimuli moved away from Category A in the direction of the fuzzy prototype.

Alternatively, the displacement of the Category A prototype away from Category B evident in Figures 5.4 through 5.6 may be accurate in the sense that it corresponds to subjects' learning. If so, this would resemble a kind of 'peak shift' in which the stimulus value associated with maximal responding is moved away from the value associated with the other response. Peak shift was originally reported for pigeons in terms of wavelength generalization gradient following discrimination training (Hanson, 1959), but is a general phenomenon, having been reported in the context of temporal and spatial discrimination for both pigeons and humans (Bizo & McMahon, 2007; Cheng & Spetch, 2002; Cheng, Spetch & Johnston, 1997), and human category learning (McLaren, Bennett, Guttman-Nahir, Kim & Mackintosh, 1995).

Another possible way in which the model might be modified or enhanced would be to change the function relating distance to Category A responding. In the current version of the model, we assumed that the odds of responding to Category A was a (negative) power function of the distance between the stimulus percept and the fuzzy prototype. This assumption was adequate to account for the present data, and in fact performed better than other monotonically-decreasing functions, such as exponential and hyperbolic, that we also investigated. However, it is possible that a different function might prove more satisfactory in the future.

In summary, the Fuzzy Prototype model is a new model for category learning, and was to account for the major pattern of results obtained in the present experiments, unlike previous models. These results included the inverted U-shape pattern for Category A and the lack of a corresponding pattern in Category B for the information integration task, and the interaction between orientation and frequency in the rule-based, frequency-relevant task. That we obtained a similar pattern of results for humans and pigeons in the information integration task strongly suggests that a common process is responsible for learning of perceptual categories in both species. The FP model explains this process in terms of learning to associate a contiguous region in the perceptual space with one category.

Relationship of Fuzzy Prototype model to existing models of category learning

The fuzzy prototype model bears some similarities to current models of category learning, in particular the striatal pattern classifier (SPC; Ashby & Waldron, 1999) and exemplar models (e.g., Nosofsky, 1986). Like the SPC, the fuzzy prototype model assumes that a region of perceptual space becomes associated with a particular response, but differs in that only one region is associated with a response, and this region is described parsimoniously as a contiguous line segment, which could correspond to a cluster of cells in the striatum. Thus the FP model imposes more constraint on the nature of the category representation than does the SPC.

Another difference between the FP model and the SPC is that the FP model assumes that subjects learn an 'A/not A' response rule during the information-integration task. This is arguably a simpler decision strategy than that assumed by the SPC, which assumes that two response categories are learned, that is, an 'A or B' strategy. The advantage of an 'A-not A' strategy is that only one prototype need be maintained in working memory. It was our intuition, based on experience with responding in the information-integration task, that the easiest way to respond accurately was to decide whether it was one particular category, and to make the other response if the stimulus did not seem to match that category. Ashby and Waldron (1999) commented that none of the participants they had debriefed had ever been able to describe the optimal rule for the task, "even when his or her performance was well described by this rule. Frequently, observers simply say that their responses were just a 'gut reaction'" (p. 366). Ashby, Ennis and Spiering (2007) have argued that these types of responses suggest that performance in the information-integration task is based on implicit learning. Thus the fuzzy prototype model provides a more specific characterization of that learning process for the information-integration tasks pioneered by Ashby and colleagues.

The FP model also shares some similarities with exemplar theory (Brooks, 1978; Nosofsky, 1986). The fuzzy prototype is represented as a set of points, which could be viewed as a set of exemplars. However, there is an important difference: According to exemplar models, subjects compare the stimulus percept on a given trial with all of the exemplars that are stored in memory. Averages are computed for both categories, and responding is determined by the relative similarity. By contrast, the FP model assumes that the subject computes the minimum distance to the set of points comprising the fuzzy prototype. This is an important distinction because the FP model emphasizes the fuzzy prototype as an integral unit, not just a collection of points.

Standardized distances (i.e., *Z* scores) were used in the computations for the FP model. As previously noted, we made this assumption to solve the problem of orientation and frequency being measured in different physical units, and because it was simpler than other alternatives, such as including separate scaling factors for each dimension. However, how this ‘units’ problem is solved is critical in terms of how the model is to be interpreted, and particularly whether it has a realistic neurobiological interpretation. If the basic assumption of the SPC is correct – that is, that there a low-resolution map of the stimulus space created in the striatum via cortical projections – then the units problem may be solved through that mapping. For example, a stimulus space of arbitrary size could be mapped onto a finite network of cells laid out as a grid by first anchoring the extreme points on each dimension to the respective ends of the grid, and then filling out the interior points through an interpolation-like process. For this to occur, cells would need to be able to respond in relativistic terms (i.e., sensitive to the distance between the percept and the extremes). In this way, the grid would be scalable and could represent, at least in principle, a two-dimensional stimulus space with arbitrary ranges on both dimensions. If the stimulus space was represented in this way, judgments of similarity in terms of Euclidean distances based on standardized values would seem plausible.

However, with this type of scalable representation there should be no effects of stimulus range. This prediction is contradicted by results from the Narrow condition in Chapter 3. In that condition, when the range of orientation values was limited to 10 degrees, accuracy decreased for both pigeons and humans, and model analyses suggested that responding was controlled solely by frequency. As currently formulated, the FP model provides no principled basis for predicting when this should occur. It may be that when trial-to-trial variation in a particular stimulus dimension is low (as it was for orientation in the Narrow condition), the process by which the stimulus space is mapped onto the striatum changes, and extreme values on a more salient dimension (i.e., frequency in the Narrow condition) become anchored exclusively in the grid. Somehow, the effective representation must then change, becoming one- rather than two-dimensional. Exactly how this happens remains unclear.

Finally, to return to one of the questions motivating the present research, what does it mean to say that subjects are ‘integrating information’? If the FP model is accurate, are subjects really ‘integrating’ different sources of information? According to Massaro and Friedman’s (1990) definition, the answer is yes. In their terms, the calculation of Euclidean distances based on standardized stimulus values requires a combination of the representations of different stimulus dimensions. We therefore conclude that pigeons as well as humans are capable of responding accurately in information-integration tasks using stimuli with dimensions that are fully separable, and that the best account of this performance – the FP model – assumes that their decision process requires the *functional equivalent* of information integration: The computation of distance in a standardized two-dimensional stimulus space. However, just how these computations are performed at the neurobiological level remains an open question. Results from this thesis suggest that in functional terms, these computations are likely to be similar for pigeons and humans, and thus any neurobiologically-based model for category learning should not be based on features that are unique to the mammalian brain.

Directions for Future Research

Although the present studies have obtained evidence that both pigeons and humans show similar deviations from optimality in the information integration and rule-based tasks which can be predicted by the FP model, much work remains to be done before this model can be established on a par with other models for category learning. Future research should explore the generality of the empirical results reported here. For example, would a similar pattern of deviation from optimality be obtained with different stimuli, such as Shepard circles (Shepard, 1964). Second, all categories were based on normal distributions. Although normal distributions have been widely used in previous research on categorization, it would be interesting to test whether similar results are obtained in other situations, such as uniform distributions. The primary characteristic of normal distributions is that stimuli become increasingly less likely to occur as the distance from the mean increases, and thus the inverted-U shaped pattern that we obtained might have occurred because stimuli at the extremes were presented infrequently. It seems reasonable to expect that subjects' learning would depend on the distributional characteristics of the stimuli; for example, with a uniform distribution the 'fuzzy prototype' may extend over a broader segment of the stimulus space than with a normal distribution. Experiments which varied the category distributions could test this prediction, and provide additional support for the FP model.

It would also be worthwhile to test the validity of the assumption of the FP model that subjects use an 'A/Not-A' response rule. This could be accomplished in a number of ways. If an A/Not-A response rule is learned, then subjects should more readily acquire a new discrimination in which the response or Gabor stimuli for Category B are changed than if those for Category A are changed. By contrast, if subjects learn an A-B response rule, there should be no systematic difference in acquisition depending on whether the stimuli for Category A or B are changed. A related experiment could be conducted in which the response associated with either Category A or B is changed after baseline training. If subjects learn an A/Not-A response rule, then it should be more difficult to learn a new discrimination in which the response for Category A is changed compared to changing the response for

Category B. Finally, although one particular category was always associated with the fuzzy prototype in the present research (i.e., Category A), it should be possible to control which category is learned through appropriate training. For example, subjects that are given pre-training with a successive discrimination in which they have to respond or not respond, depending on whether a stimulus is a member of Category B should be more likely to apply a B/Not-B response rule when subsequently trained in the A vs B discrimination. Experiments like these could provide additional support for the FP model, or else show that an alternative account is necessary.

Additional empirical research is also needed, particularly to test the generality of our finding that humans and pigeons apparently use a common process for visual category learning. For example, we found that responding of pigeons in the RB-frequency relevant condition showed evidence of an interaction between frequency and orientation, such that pigeons were more accurate when the orientation was near vertical than when it was near horizontal. Although reasons for this finding are unclear, it is important to test whether it can also be obtained with humans. Given the strong evidence for two category learning systems in humans, one based on reinforcement learning (i.e., trial and error) used in the information integration task and another based on explicit hypothesis testing and used in the rule-based task (Maddox & Ashby, 2003), the similarity between the pigeon and human results in Chapter 3 suggests that both species may use a common system for reinforcement learning. It is unlikely that pigeons would use a separate category learning system for the rule-based and information-integration tasks, for not only *a priori* reasons, but also the fact that accuracy levels and acquisition rates for the two tasks were not systematically different in Chapter 2. Thus if humans failed to show an interaction between orientation and frequency in the RB-frequency relevant task, the difference with the pigeon results might result from humans using an explicit system.

The results of the RB condition in Experiment 1 of Chapter 2 also suggest the possibility of a perceptual interaction between the orientation and frequency dimensions. For example, based on pigeons' results, it could be concluded that discrimination of frequency was more accurate when the orientations were closer to zero degrees (horizontal) and less accurate when orientations were closer to 90 degrees (vertical). The raises an important question about the separability of the dimensions used in Gabor patches. In order for stimuli to be fully separable they must be analyzable in terms of their component dimensions, and the effects of those dimensions should be independent. Although the frequency and orientation of Gabor patches have been taken as fully separable by previous researchers (e.g., Maddox & Ashby 2004, Maddox, Ashby & Bohil 2003), our results may call that into question. A logical next step to test this issue of separability would be to conduct a follow up RB experiment with orientation as the relevant dimension and a relatively narrow range of stimuli with pigeons and also a human experiment comparing the RB orientation and RB frequency condition's performance.

Gabor patches are artificial, computer-generated stimuli and thus their ecological validity might be questioned. However, whether inferences based on research using Gabor stimuli might be limited due to this fact has not been given much consideration. Anecdotal evidence suggests that Gabor patches as used in the II human condition in Chapter 2 required a non-verbalizable strategy. In other words when participants were asked about their strategy for accuracy they reported that it was difficult to describe or that they figured it out "through experience". This is consistent with everyday human experience when solving multidimensional problems that in spite of our ability to achieve high accuracy we often can't describe how we do it. In this way Gabor patches in the II condition arguably mimic more natural categorization tasks where learning through experience is required. A future experiment that uses naturalistic yet separable stimuli would be interesting to conduct. Then the differences and similarities could be analyzed and shed further light on the issue of ecological validity with the II and RB tasks.

The question of absolute versus relative stimulus values is also relevant. The current experiment manipulations all involved orientation values that were centered on 45 degrees and ranged from no less than 10 degrees (from horizontal) to 80 degrees. Based on the results of Experiment 1 it was clear that accuracy was not only higher near the extremes of the orientation values but also when orientation was closer to horizontal. Thus a future experiment using a range of orientation values centered on a value other than 45 degrees would be a logical next step. This would help determine whether the accuracy patterns were due to the range of stimulus values for orientation and specifically whether Gabor orientations closer to vertical are more difficult to categorize.

Conclusion

Empirical research in this thesis has shown that learning of visual categories by pigeons and humans in ‘information integration’ category tasks shows intriguing similarities. These are category structures which cannot be described in terms of simple verbal rules, and are learned through the basis of trial-and-error experience – reinforcement learning. We found that performances of pigeons and humans deviated from optimality – and predictions of current models such as the GLC – in similar and systematic ways. This result strongly suggests the presence of a single process for visual category learning, common across species. We have proposed a ‘fuzzy prototype’ model for this process, which assumes that subjects learn to associate a line segment in the stimulus space with one particular response. This model provided a parsimonious account of our results, and performed better than the GLC or other current models for category learning. Although whether the assumptions of the model will be supported by future experiments is unclear, we believe that researchers should continue to use the comparative paradigm employed in this thesis, to better understand how organisms learn visual categories on the basis of experience.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *16*(6), 716-723.

Alsop, B. (2004). Signal-detection analyses of conditional discrimination and delayed matching-to-sample performance. *Journal of the Experimental Analysis of Behavior, 82*(1), 57-69.

Alsop, B. (1998). Receiver operating characteristics from nonhuman animals: Some implications and directions for research with humans. *Psychonomic Bulletin and Review, 5*(3), 239-252.

Alsop, B. (1991). Behavioral models of signal detection and detection models of choice. In M. L. Commons, J. A. Nevin, & M. C. Davison (Eds.), *Signal detection: Mechanisms, models, and applications.* Hillsdale, NJ: Erlbaum.

Ashby, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics, 44*(3), 195-204.

Ashby, F. G. (Ed.). (1992). *Multidimensional Models of Perception and Cognition*. Hillsdale, New Jersey: Lawrence Erlbum Associates.

Ashby, F. G., Alfonso-Reese, L. A., Turken, & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442-481.

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, *114*, 632-656.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*(1), 33-53.

Ashby, F. G., Ell, S. W., & Waldron, E. M. (2003). Procedural learning in perceptual categorization. *Memory & Cognition, 31*(7), 1114-1125.

Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception & Performance, 16*(3), 598-612.

Ashby, F. G. & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception & Performance, 18*(1), 50-71.

Ashby, F. G., & Maddox, W. T. (1998). *Stimulus categorization*. San Diego, CA: Academic Press.

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Reviews of Psychology, 56*.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition, 30*(5), 666-677.

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception &*

Psychophysics, 61(6), 1178-1199.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*(2), 154-179.

Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review, 6*(3), 363-378.

Avriel, M. (2003). *Nonlinear Programming: Analysis and Methods.* Mineola, *NY*: Dover Publishing.

Balsam, P. D. (1988) Selection, representation and equialence of controlling stimuli. In R. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.) *Handbook of experimental psychology*. New York, NY: Wiley.

Berg, M. E., Grace, R. C., & Cook, R. G. (2009). Categorization of Multidimensional Stimuli by Pigeons. *In Preparation*.

Bizo, L. A., & McMahon, C. V. (2007). Temporal generalization and peak shift in humans. *Learning & Behavior, 35*(2), 123-130.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.) *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum.

Cerella J. (1990). Pigeon pattern perception: limits on perspective invariance. *Perception 19*(2), 141–159

Cerella, J. (1977). Absence of perspective processing in the pigeon. *Pattern Recognition*, *9*(1), 65-68

Cheng, K., & Spetch, M. L. (2002). Spatial generalization and peak shift in humans. *Learning & Motivation, 33*(4), 358-389.

Cheng, K., Spetch, M. L., & Johnston, M. (1997). Spatial peak shift and generalization in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 23*(4), 469-481.

Cook, R. G., & Smith, J. D. (2006). Stages of Abstraction and Exemplar Memorization in Pigeon Category Learning. *Psychological Science 17*(12), 1059-1067.

Cook, R. G., Katz, J. S., & Cavoto, K. K. (1997). Pigeon same-different concept learning with multiple stimulus classes. *Journal of Experimental Psychology: Animal Behavior Processes, 23*(4): 417-433.

Cook, R. G., Katz, J. S., & Kelly, D. M. (1999). Pictorial same-different categorical learning and discrimination in pigeons. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 18*(5-6): 805-843.

Davenport, R. K., & Rogers, C. M. (1971). Perception of photographs by apes. *Behaviour, 39*(2-4), 318-320.

Davison, M., & McCarthy, D. (1989). Effects of relative reinforcer frequency on complex color detection. *Journal of the Experimental Analysis of Behavior, 51*(3), 291-315.

Davison, M., & Nevin, J. A. (1999). Stimuli, reinforcers, and behavior: An integration. *Journal of the Experimental Analysis of Behavior, 71*(3), 439-482.

Davison, M. C., & Tustin, R. D. (1978). The relation between the generalized matching law and signal-detection theory. *Journal of the Experimental Analysis of*

Behavior, 29(2), 331-336.

Emmerton J. (1986). The pigeon's discrimination of movement patterns (Lissajous figures) and contour-dependent rotational invariance. *Perception, 15*(5), 573–588.

Frank, A. J., & Wasserman, E. A. (2005). Associative symmetry in the pigeon after successive matching-to-sample training. *Journal of the Experimental Analysis of Behavior, 84*(2), 147-165.

Goffe, W. L., Ferrier, G. D., & Rogers, J. (1994). Global Optimization of Statistical Functions with Simulated Annealing. *Journal of Econometrics, 60*(1), 65-99.

Gravetter F., & Lockhead, G.R. (1973). Criterial range as a frame of reference for stimulus judgment. *Psychological Review, 80*(3), 203-16.

Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology, 58*(3), 321-334.

Hayes, S. C. (1989). Nonhumans have not yet shown stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 51*(3), 385-392.

Herbranson, W. T., Fremouw, T., & Shimp, C. P. (1999). The randomization procedure in the study of categorization of multidimensional stimuli by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 25*(1), 113-134.

Herbranson, W. T., Fremouw, T., & Shimp, C. P. (2002). Categorization a moving target in terms of its speed, direction, or both. *Journal of the Experimental Analysis of Behavior,* 78(3): 249-270.

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon.

Science, 146 (Whole No. 3643), 549-550.

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 2*(4), 285-302.

Hinson, J. M., & Lockhead, G. R. (1986). Range effects in successive discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 12*(3): 270-276.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65*(1), 185-241, 341-153.

Huber, L. (2001). Visual categorization in pigeons. In R. G. Cook, (Ed.), *Avian visual cognition* [On-line]. Available: www.pigeon.psy.tufts.edu/avc/huber/

Jitsumori, M., & Makino, H. (2004). Recognition of static and dynamic images of depth-rotated human faces by pigeons. *Learning & Behavior, 32*(2): 145-165.

Jitsumori, M., Sieman, M., Lehr, M., & Delius, J. D. (2002). A new approach to the formation of equivalence classes in pigeons. *Journal of the Experimental Analysis of Behavior, 78*(3): 397-408.

Kastak, C. R., Schusterman, R. J., & Kastak, D. (2001). Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior, 76*(2), 131-158.

Krantz, D. H., & Tversky, A. (1975). Similarity of rectangles: An analysis of subjective dimensions. *Journal of Mathematical Psychology, 12*(1), 4-34.

Lazareva, O. F., Smirnova, A. A., Bagozkaja, M. S., Rayevsky, V. V., Wasserman, E. A., & Zorina, Z. A. (2004). Transitive responding in hooded crows requires linearly ordered stimuli. *Journal of the Experimental Analysis of Behavior*, *82*(1), 1-19.

Lea, S. E. G., & Dittrich W. H. (2000). *What do birds see in moving video images?* New York, NY, US, Psychology Press.

Lionello-DeNolf, K. M. (2009). The search for symmetry: 25 years in review. *Learning & Behavior, 37*(3), 188-203.

Lionello-DeNolf, K. M., & Urcuioli, P. J. (2002). Stimulus control topographies and tests of symmetry in pigeons. *Journal of the Experimental Analysis of Behavior, 78*(3), 467-495.

Lockhead, G. R. (2004). Absolute judgements are relative: A reinterpretation of Some Psychophysical ideas. *Review of General Psychology,* 8(4), 265-272.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A Network Model of Category Learning. *Psychological Review, 111*(2), 309-332.

Machado, A. & Cevik, M. (1997). The discrimination of relative frequency by pigeons. *Journal of the Experimental Analysis of Behavior***,** 67(1), 11-41.

Macmillan, N.A., & Creelman, C.D. (2005). *Detection theory: A user's guide*. (2nd Edition). Mahwah, NJ: Erlbaum

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics, 53*(1), 49-70.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes, 66*(3), 309-332.

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*(4), 650-662.

Maddox, W. T., Ashby, F. G., & Gottlob, L. R. (1998). Response time distributions in multidimensional perceptual categorization. *Perception & Psychophysics, 60*(4), 620-637.

Maddox, W. T., Ashby, F. G., Ing, A. D. & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition, 32*(4), 582-591.

Maddox, W. T., & Filoteo, J. V. (2001). Striatal contributions to category learning: Quantitative modeling of simple linear and complex nonlinear rule learning in patients with Parkinson's disease. *Journal of the International Neuropsychological Society, 7*(6), 710-727.

Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category Number Impacts Rule-Based but Not Information-Integration Category Learning: Further Evidence for Dissociable Category-Learning Systems. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*(1), 227-245.

Malott, R. W., & Siddall, J. D. (1972). Acquisition of the people concept in pigeons. *Psychological Reports, 31*(1), 3-13.

Marcum, J. (1960). A statistical theory of target detection by pulsed radar. *IRE Transactions on Information Theory, 6*(2), 59-267.

Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review, 97*(2), 225-252.

McLaren, I. P. L., Bennett, C. H., Guttman-Nahir, T., Kim, K., & Mackintosh, N. J. (1995). Prototype effects and peak shift in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 662-673.

Miller, G.A. (1956) The magical number seven, plus or minus two. *Psychological Review, 63*(1), 81–97.

Miller, R. R., & Barnet, R. C. (1993). The role of time in elementary associations. *Current Directions in Psychological Science*, *2*(1), 106-111.

Neiworth, J. J., & Rilling, M. E. (1987). A method for studying imagery in animals. *Journal of Experimental Psychology: Animal Behavior Processes, 13*(3), 203-214.

Neiworth, J. J., & Wright, A. A. (1994). Monkeys (Macaca mulatta) learn category matching in a nonidentical same-different task. *Journal of Experimental Psychology: Animal Behavior Processes***,** *20*(4), 429-435.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., & McKinley, S. C. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition***,** *22*(3), 352-369.

Parducci, A., & Wedell, D. H. (1986). The category effect with rating scales: Number of categories, number of stimuli, and method of presentation. *Journal of Experimental Psychology: Human Perception and Performance, 12*(4), 496-516.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(31), 353-363.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism-Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, *63*(5), 737-748.

Saint-Cyr, J. R., Ungerleider, L. G., & Desimone, R. (1990). Organization of visual cortical inputs to the striatum and subsequent outputs to the pallido-nigral complex in the monkey. *The Journal of Comparative Neurology, 298*(2), 129-156.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1*(1), 54-87.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*(13, Whole No. 517), 42.

Sherwonit, B., Schultz, J. (1991). *Iditarod: The Great Race to Nome*. Seattle, WA, US, Sasquatch Books.

Sidman, M. (1990). *Equivalence relations: Where do they come from?* Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Sidman, M. (1992). *Equivalence relations: Some basic considerations*. Reno, NV: Context Press.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior*, 37(1), 23-44.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37(1), 5-22.

Siegel, R. K., & Honig, W. K. (1970). Pigeon concept formation: Successive and simultaneous acquisition. *Journal of the Experimental Analysis of Behavior, 13*(3), 385-390.

Skinner, B. F. (1960). Pigeons in a pelican. *American Psychologist, 15,* 28-37.

Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category Learning in Rhesus Monkeys: A Study of the Shepard, Hovland, and Jenkins (1961) Tasks. *Journal of Experimental Psychology: General, 133*(3), 398-414.

Spetch, M. L., Kelly, D. M., & Reid, S. (1999). Recognition of objects and spatial relations in pictures across changes in viewpoint. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition***,** 18(5-6), 729-764.

Spetch, M. L., Kelly, D. M., & Reid, S. (2000). *Recognition of objects and spatial relations in pictures across changes in viewpoint.* New York, NY, US, Psychology Press.

Swets, J. A. (1964). Signal detection and recognition by human observers: Contemporary readings, Hoboken, NJ: Wiley.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review 34*(3), 273-286.

Urcuioli, P. J. (2008). Associative symmetry, antissymmetry, and a theory of pigeons' equivalence-class formation. *Journal of the Experimental Analysis of Behavior, 90*(3), 257-282.

Vaughan, W., Jr. (1988). Formation of equivalence sets in pigeons. *Animal Behavior Processes,* 14, 36-42.

Wasserman, E. A., J. L. Gagliardi, J. L., Cook, B. R., Kirkpatrick-Steger, K., Astley, S. L., & Biederman, I. (1996). The pigeon's recognition of drawings of depth-rotated stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 22*(2), 205-221.

Wasserman, E. A., J. A. Hugart, J. A., & Kirkpatrick-Steger, K. (1995). Pigeons show same different conceptualization after training with complex visual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *21*(3), 248-252.

Wasserman, E. A.,Young, M. E., & Nolan, B. C. (2000). Display variability and spatial organization as contributors to the pigeon's discrimination of complex visual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 26*(2), 133-143.

Walraff, H. G. (2005). Avian Navigation: Pigeon Homing as a Paradigm, Berlin: Springer.

Watanabe, S. (1993). Object picture equivalence in the pigeon: An analysis with natural concept and pseudoconcept discriminations. *Behavioural Processes, 30*(3), 225-231.

Watanabe, S. (1997). Visual discrimination of real objects and pictures in pigeons. *Animal Learning & Behavior, 25*(2), 185-192.

Watanabe, S. (2001). Van Gogh, Chagall and pigeons: Picture discrimination in pigeons and humans. *Animal Cognition, 4*(3-4), 147-151.

Watanabe, S., Sakamoto, J., & Wakita, M. (1995). Pigeons' discrimination of painting by Monet and Picasso. *Journal of the Experimental Analysis of Behavior, 63*(2), 165-174.

White, K. G., & Wixted, J. T. (1999). Psychophysics of remembering. *Journal of the Experimental Analysis of Behavior, 71*(1), 91-113.

Yao, J., Krolak, P., & Steele, C. (1995). The Generalized Gabor Transform. *IEEE Transactions on Image Processing, 4*(7), 978-988.

Young, M. E., Wasserman, E. A., & Dalrymple, R. M. (1997). Effects of number of items on the pigeon's discrimination of same from different visual displays. *Journal of Experimental Psychology: Animal Behavior Processes, 23*(4), 491-501.

Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior, 78*(3), 237-248.

TECHNICAL APPENDIX

Our plan to conduct comparative research on categorization with pigeons and humans required a software and hardware system with special requirements, including the ability to generate and display the Gabor images. After displaying the images, the system needed to record and evaluate inputs via touch screen and response keys (for pigeons), or via the keyboard (for humans). For experiments with pigeons, a touch-enabled screen was necessary for the orienting responses required after the display of the images. After each response the system needed to provide feedback, either in the form of grain or on-screen feedback for pigeons and humans, respectively, based on the categorization response. The system also had to record the location and time stamp for each response.

Apart from the core functionality requirements there were some secondary goals for the hardware/software system. Rather than display pre-generated Jpeg images of the Gabor patches, we designed a system that could generate the images in real time. This allowed for changes in the display stimuli to be made with ease and with precision and for greater flexibility in the experimental procedures. We also wanted a system that could be easily run from a laptop computer to facilitate collection of data with human participants. By having the same control file and system for both pigeons and humans we eliminated the possible errors associated with having different programs. Thus the system represents a platform for comparative research on categorization with pigeons and humans and should be useful for studies in the future as well.

New System - Hardware

We began the project by designing a pigeon chamber that would include the necessary features to support the research, including a touchscreen display, response keys and grain feeder. These devices needed to be controlled from a single computer, and thus to provide a common interface we used a readily-available I/O device based on the Universal Serial Bus (USB) port. The advantage of such an interface is that it is inexpensive, available on both laptop and desktop computers, and multiple interfaces can be controlled at one time by one computer system. Such devices are available from a variety of sources, but we selected the USB-1024LS manufactured by Measurement Computing (http://www.measurementcomputing.com/), which is shown in Figure 1.

Figure 1. PMD-1024LS DIO Interface

The PMD-1024LS has 24 bits of digital input/output (DIO), is based on industry standard 82C55 control circuitry and is relatively inexpensive (USD$99.00). The size is small meeting our portability requirements and the hardware control can be achieved through the use of any major programming language including C, C++, Java, Python and many others. The unit also ships with code and control library code samples. Each of the units can be uniquely identified which gives the user the ability to interface multiple devices to a single computer and control each one independently.

We then searched for a visual display with integrated touchscreen. We looked for a small touch screen with high resolution for detailed image display and compact screen size in order to permit a grain feeder to be mounted on the front panel of the pigeon chamber for reinforcer delivery. Based on those requirements we choose the Open Frame Touch monitor MTF064D available from TouchScreens.com. This touch screen measures 6.4” diagonally and includes an Elo Systems AccuTouch 5-Wire Resistive touch overlay with USB interface. The screen itself uses a standard VGA connector and displays true VGA 640 (H) x 480(V) resolution. The monitor is shown in Figure 2.

Figure 2. LCD Touch screen monitors

The upper panel of Figure 3 shows a schematic of the intelligence panel, which was mounted on the front of the experimental chambers. There were four response keys mounted near the corners of the touchscreen, which could be illuminated with five different colors (red, green, yellow, white and blue) via rear-mounted LEDs. There also was a houselight located above the touchscreen. A picture of the completed system (with pigeon) is shown in the lower panel.

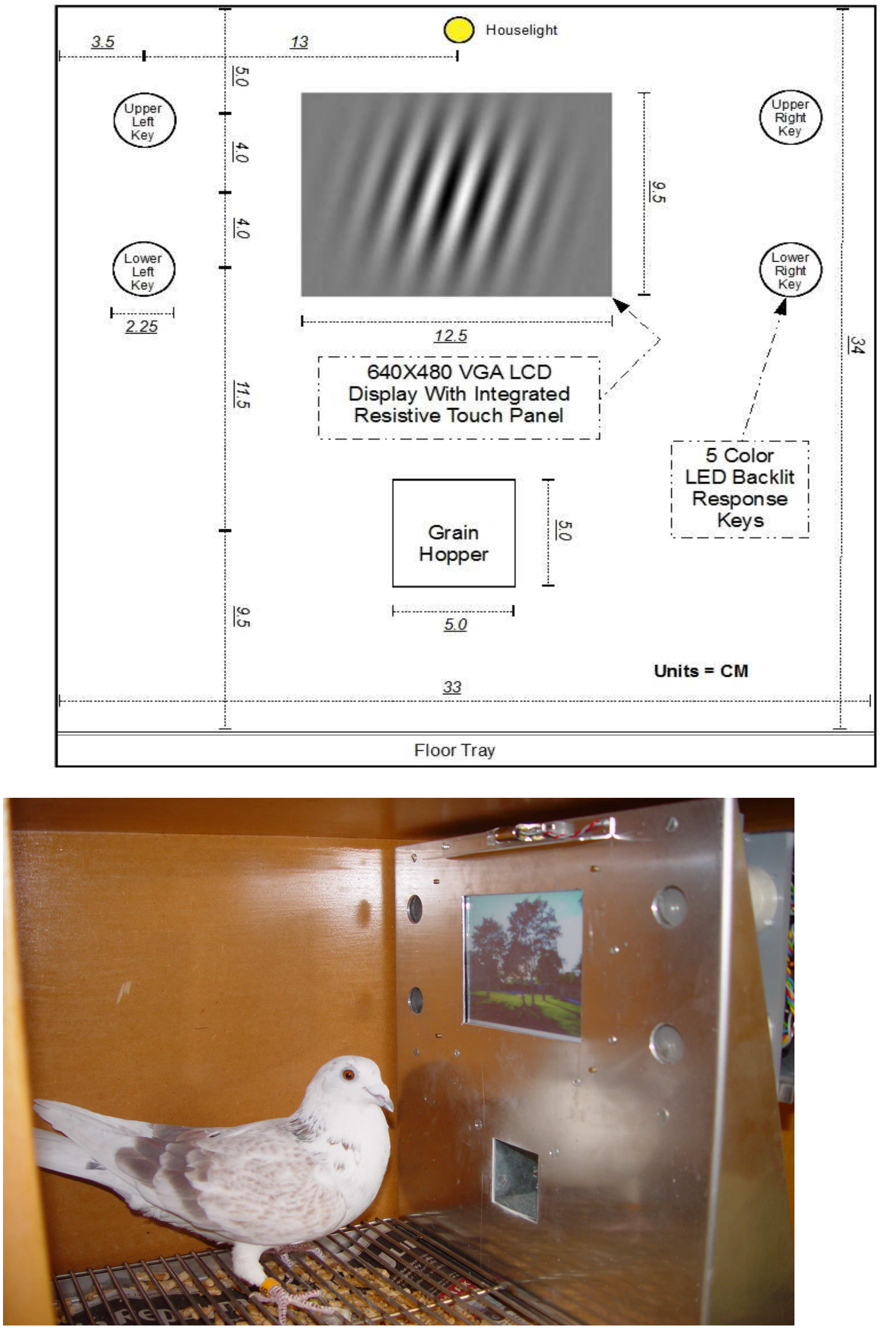


Figure 3. The top panel shows a schematic drawing of the intelligence panel design. The bottom panel shows a drawing of the finished product

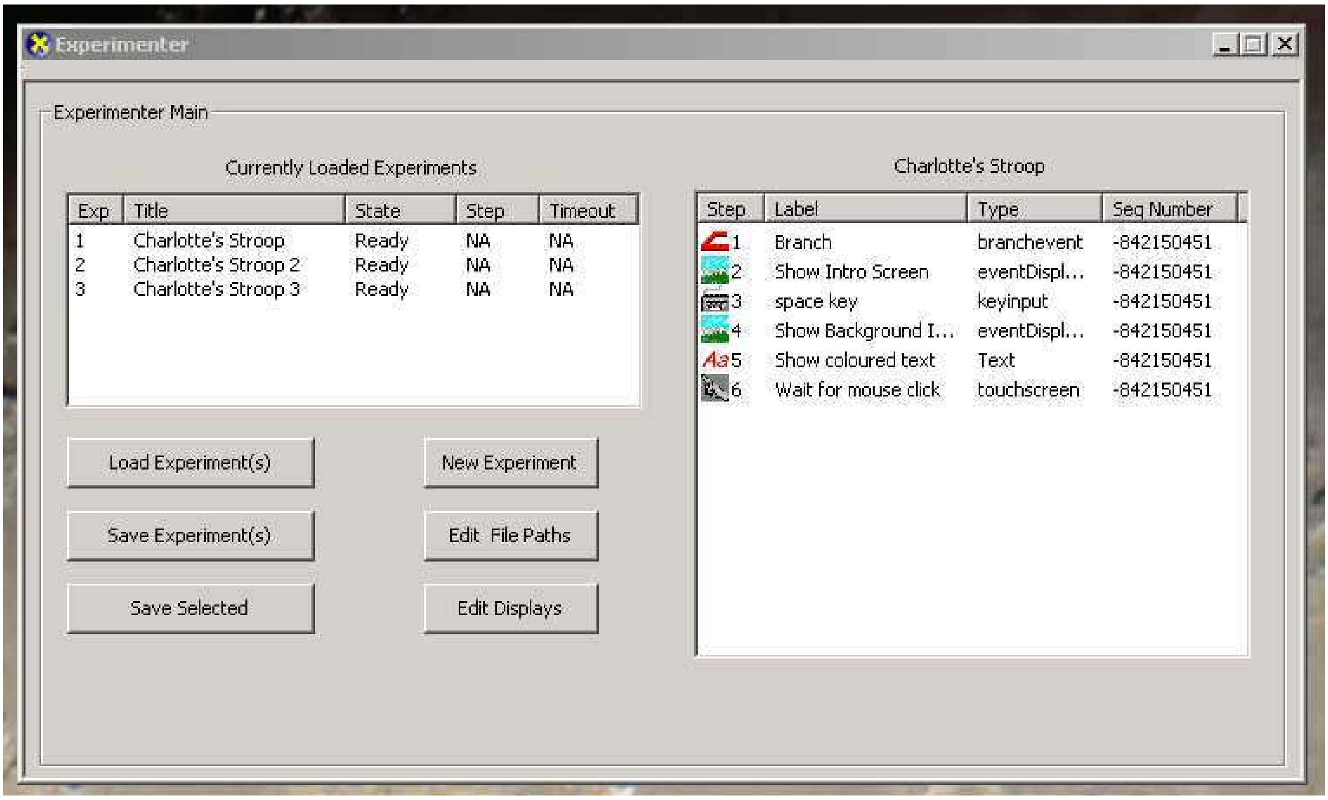
In total, four experimental chambers identical to that shown in Figure 25 could be controlled by a single computer. The computer specifications included the following: Microsoft Windows XP operating system, 2.4 gigahertz American Micro Devices (AMD) processor, 2 gigabytes of ram, and 5 VGA video outputs (4 PCI video cards and 1 AGP card).

New System - Software

The software required to run the system, control the experiments and collect the data was developed after the hardware specifications had been finalized. Our goal was to design a control system that could run all of the hardware concurrently with at least 1-millisecond accuracy, including the four touch screen interfaces, VGA outputs and I/O interfaces, and to provide a high-level utility that made it easy to program experiments without the need for the operator to learn a specialized language. Once the goals were defined, we began developing the software in collaboration with the departmental programmer.

The resulting software is called *Experimenter* and has turned out to be a useful system not only for research with pigeons, but also for laboratory coursework and research conducted at the Van De Veer Institute for brain research (Christchurch Medical School) using eye-tracking hardware. The software was written in Visual C++ and allows for display of file-based images, but also includes a utility that generates the Gabor patches on the fly based on a predefined list of frequency and orientation values.

The list can be randomized with a common seed value so that each day the pigeons receive the Gabor stimuli in a unique random ordering that is common for all of the subjects running in the session because it is based on a common seed value. Other elements of the experiment are created or changed using a simple graphical step-by-step interface that is menu-driven and graphic intensive. Experimental steps can be easily sequenced based on text files and the steps themselves can each be randomized in with or without replacement fashion. Steps can also be made contingent on specific hardware inputs and the outputs themselves can also be sequenced randomly or based on a predefined list. Figure 4 shows a screen shot example of some of the menus available in *Experimenter*.

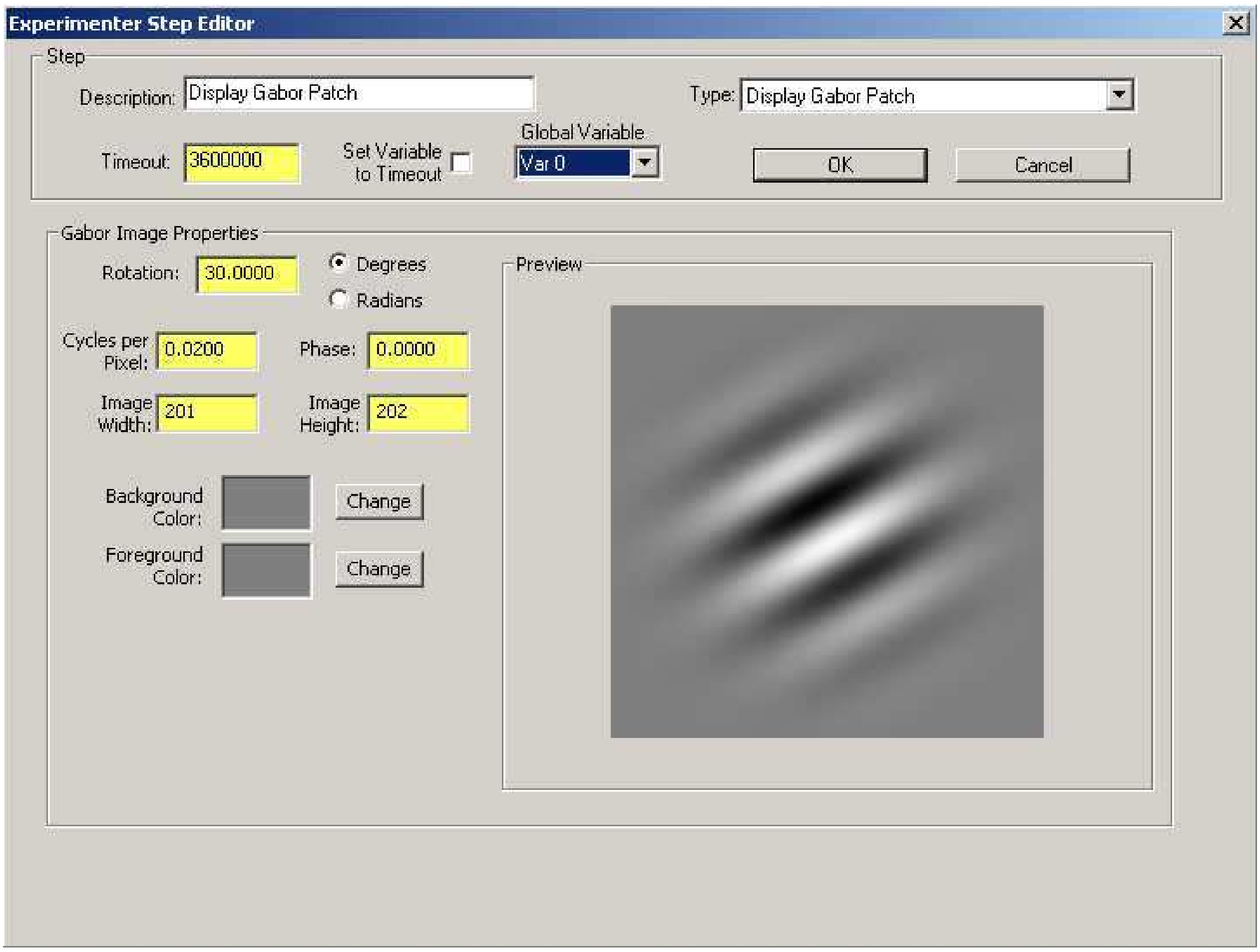


Currently Loaded Experiments

Exp	Title	State	Step	Timeout
1	Charlotte's Stroop	Ready	NA	NA
2	Charlotte's Stroop 2	Ready	NA	NA
3	Charlotte's Stroop 3	Ready	NA	NA

Charlotte's Stroop

Step	Label	Type	Seq Number
1	Branch	branchevent	-842150451
2	Show Intro Screen	eventDispl...	-842150451
3	space key	keyinput	-842150451
4	Show Background I...	eventDispl...	-842150451
5	Show coloured text	Text	-842150451
6	Wait for mouse click	touchscreen	-842150451



Experimenter Step Editor

Step

Description: Display Gabor Patch

Type: Display Gabor Patch

Timeout: 3600000

Set Variable to Timeout ☐

Global Variable: Var 0

Gabor Image Properties

Rotation: 30.0000

● Degrees

○ Radians

Cycles per Pixel: 0.0200

Phase: 0.0000

Image Width: 201

Image Height: 202

Background Color:

Foreground Color:

Figure 4. Sample menus available in the software

The top panel of Figure 4 shows the main *Experimenter* control window. In this window, specific control programs (i.e., experiments) can be loaded to run, or edited. While an experiment is running, progress can be monitored and any problems noted. The bottom panel shows a screen capture of the tool for generating Gabor stimuli. Various characteristics of the stimuli can be specified, including size and color, but for our purposes the two most important features are frequency and orientation, which are defined as ‘cycles per pixel’ and ‘rotation’, respectively, and can be changed by right-clicking the appropriate boxes.

The enclosed CD contains a fully functional version of the *Experimenter* software package, as well as an example of the software used to run the human experiment contained in Chapter 3. Appendix A contains instructions for running the experiment as a demonstration on any computer running the Windows XP operating system. Appendix B is a user’s manual, which describes the important features of *Experimenter*.

Appendix A

Place the enclosed CD into any Windows XP enabled computer and open the 'Open First' text file. Install *Experimenter* with all of the default options by hitting the 'Next' button until installation is complete.

Now the necessary demonstration files must be copied into the *Experimenter* folder in order to run the program. This is done by copying and replacing the folders marked 'Images' and 'Experiments' from the enclosed CD to the same named default folders that were created when installing *Experimenter*.exe. Open Windows explorer to the folder entitled '*Experimenter*', which is usually C:*Experimenter* on most computers. Then completely replace the folders entitled 'Images' and 'Experiments' with the ones from the CD. Now the demonstration program is ready to run.

To run the demonstration program, first launch the *Experimenter* program by double clicking on the newly created desktop shortcut. *Experimenter* will now open to the main interface window. Next open the demonstration configuration file by clicking on the 'Load Experiment' button. After clicking the button, an explorer window will open allowing you to browse and locate the demonstration program entitled 'Human1.xml'. Once that has been selected, it may be run by right clicking on it in the experiments list. A right click gives several options for running the procedure; select the 'initialize and run' option from the list. Next, a data recording option window will open, please leave all options in this window with the default settings and click the 'save' button. The demonstration experiment will now begin and the on-screen instructions should be followed. Categorization responses are made by pressing either the A or L keys on the keyboard. To exit the program, press the Escape key on the keyboard.

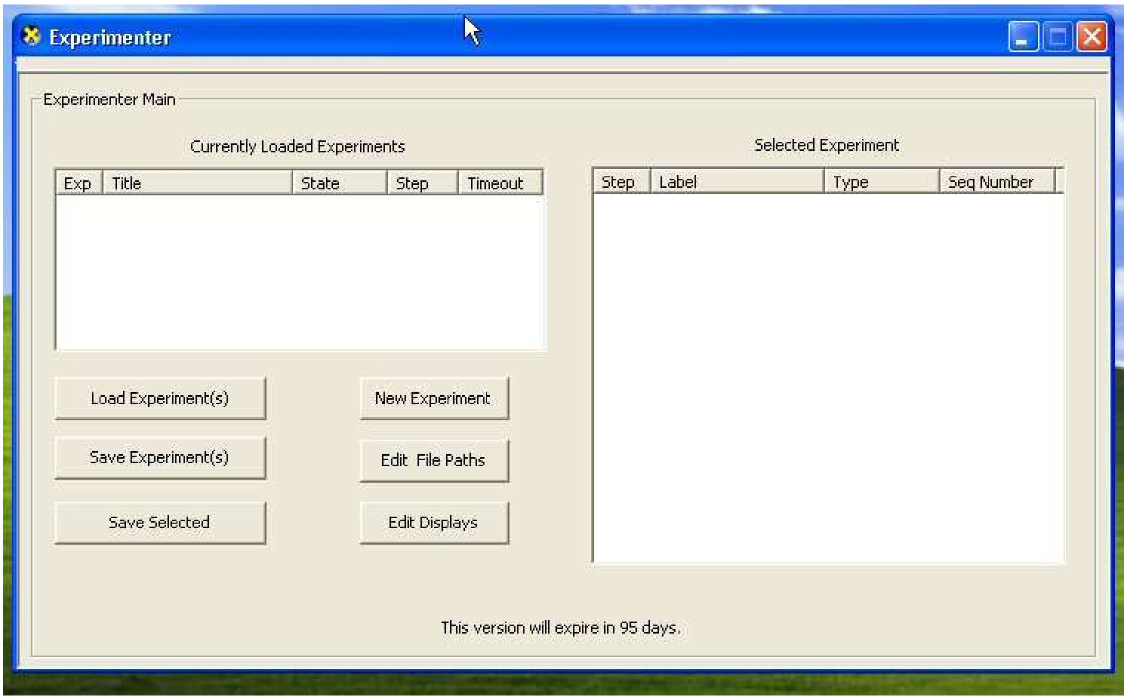
Additional installation, running and editing instructions are found in the comprehensive instruction manual for *Experimenter* found in Appendix B. The instruction manual was prepared to make the program accessible to other users who had little or no experience with the system but wanted to use it for research or teaching purposes.

Appendix B

Experimenter Overview

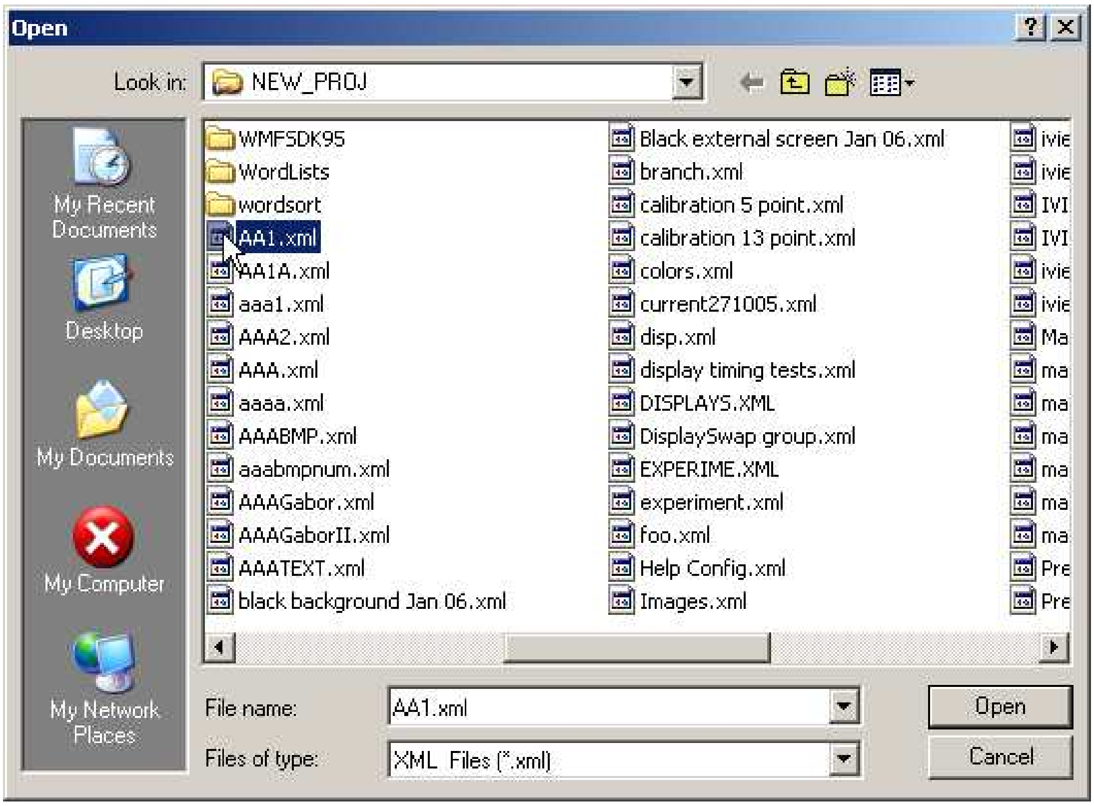
Experimenter Main Screen

When *Experimenter* is first run the Main Screen appears:



This version will expire in 95 days.

The first thing we are going to do is to load an existing configuration by clicking the ‘Load Experiment(s) button. This configuration was previously created in *Experimenter* and then saved. This is the configuration load screen:



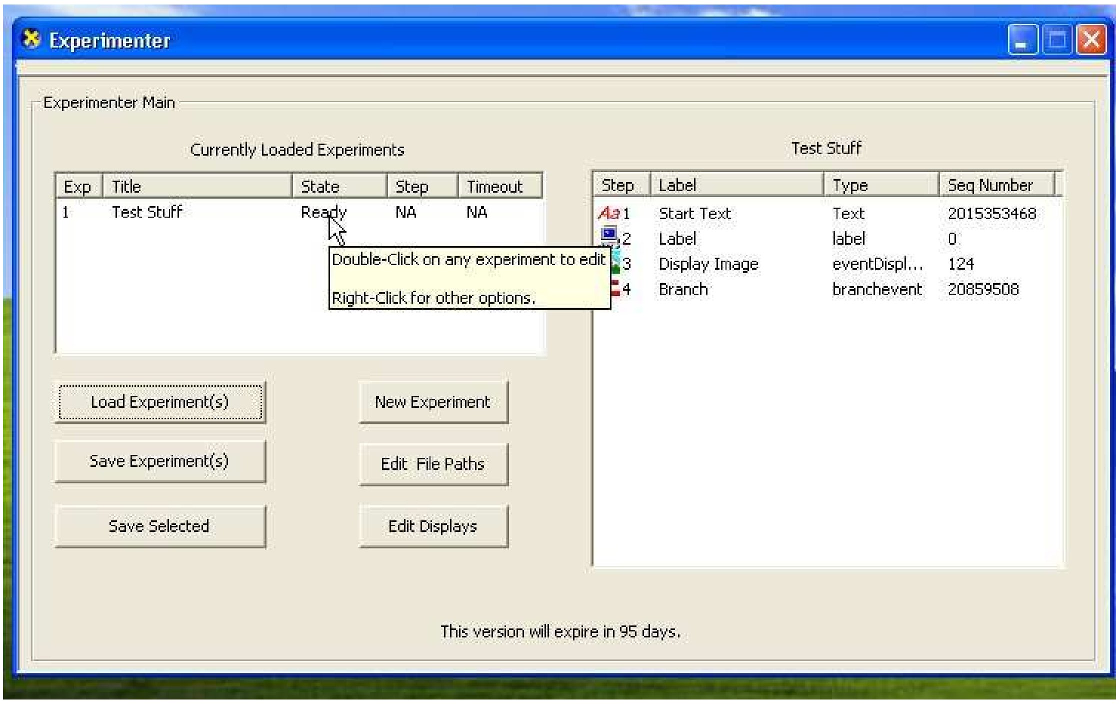
Look in: NEW_PROJ

File name: AA1.xml

Files of type: XML Files (*.xml)

We will select the configuration file, AA1.xml. Note that all Experiment Configuration files are in Extensible Markup Language (XML). This is a human-readable text file format which is widely used on the Internet for platform-independent data exchange. This is not the most efficient way to represent data because of its large file size, but is very portable, is ubiquitous on the Internet, can be displayed in any web browser and in a pinch can be manually edited using any text editor.

This is the Main Screen with the configuration loaded:



Currently Loaded Experiments

Exp	Title	State	Step	Timeout
1	Test Stuff	Ready	NA	NA

Double-Click on any experiment to edit

Right-Click for other options.

Test Stuff

Step	Label	Type	Seq Number
1	Start Text	Text	2015353468
2	Label	label	0
3	Display Image	eventDispl...	124
4	Branch	branchevent	20859508

This version will expire in 95 days.

This configuration consists of just one experiment. It could just as easily have 2 or 10 or 1000 experiments, all of which would be listed in the left pane. The right pane lists the steps in the experiment which is currently selected in the left hand pane. In this case there is only the single experiment so only its steps are listed.

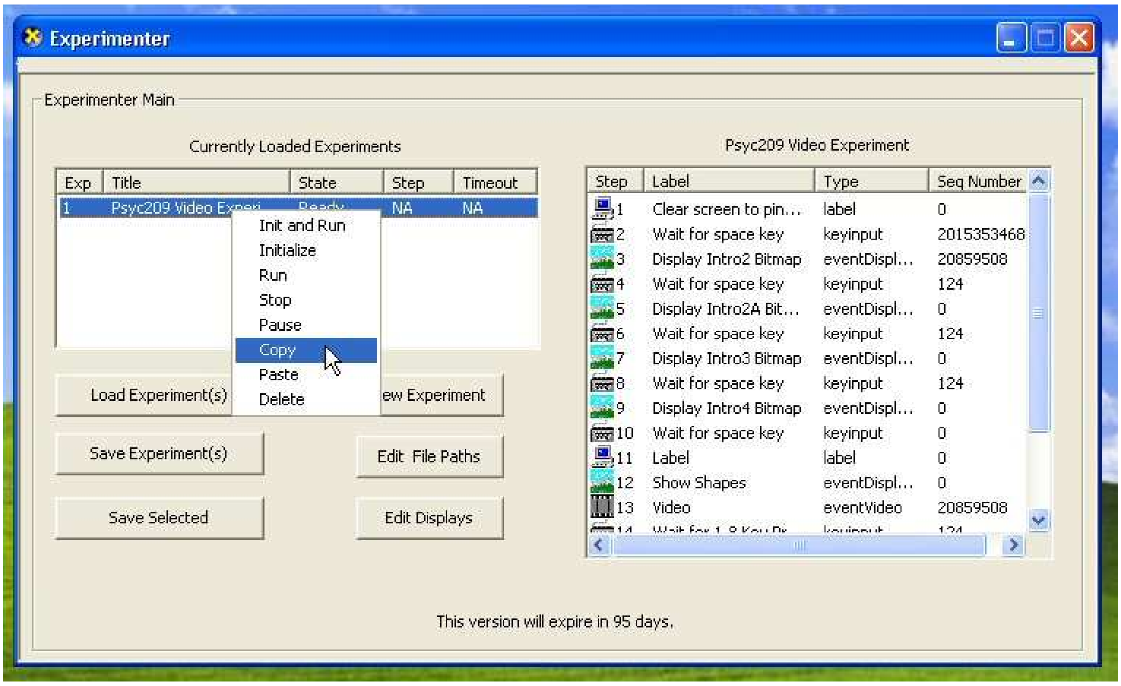
The only items of interest in the left pane right now are the Experiment Number, Title and State. There are currently four possible States as follows:

- **Ready –** this is the default state and indicates that the experiment is ready for editing or initialization preparatory to running the experiment.
- **Initialized –** before running the experiment it must be ‘initialized’. Once the experiment has been initialized it will be in this state. More on this later.
- **Running** – This indicates that the experiment is executing and collecting data.
- **Finished** – After the experiment has completed it enters this state showing that all steps have been executed.

Copying Experiments

It is frequently useful to create multiple copies of experiments with minor variations between each experiment. In *Experimenter* this can be easily done by first creating an experiment or loading an existing configuration with the experiment in it, making a copy of the experiment and then editing the copy to make the changes. This can save a lot of time while minimizing copying errors. Just remember that errors in the original will propagate!

To copy an experiment in the Main Screen, simply Right-Click on the experiment to be copied and select ‘Copy’:



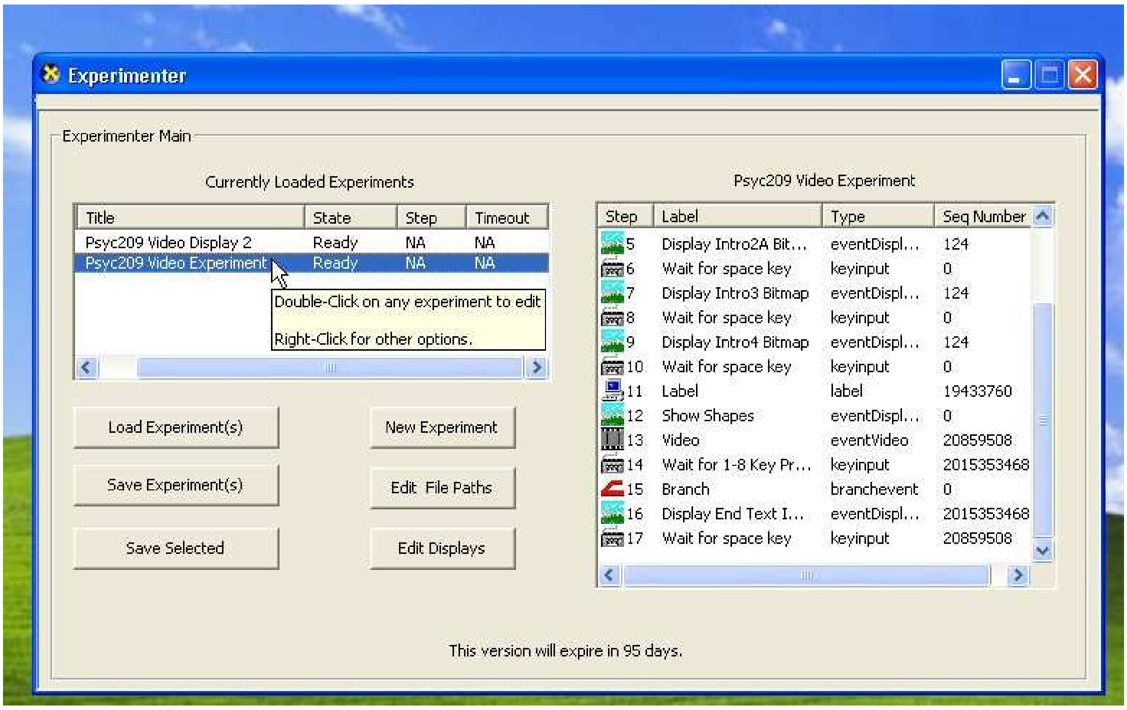
Exp	Title	State	Step	Timeout
1	Psyc209 Video Experi	Ready	NA	NA

Step	Label	Type	Seq Number
1	Clear screen to pin…	label	0
2	Wait for space key	keyinput	2015353468
3	Display Intro2 Bitmap	eventDispl…	20859508
4	Wait for space key	keyinput	124
5	Display Intro2A Bit…	eventDispl…	0
6	Wait for space key	keyinput	124
7	Display Intro3 Bitmap	eventDispl…	0
8	Wait for space key	keyinput	124
9	Display Intro4 Bitmap	eventDispl…	0
10	Wait for space key	keyinput	0
11	Label	label	0
12	Show Shapes	eventDispl…	0
13	Video	eventVideo	20859508

This version will expire in 95 days.

Now right-click anywhere in the left-pane and select ‘Paste’:

You now have a duplicate copy of the first experiment which you can now edit to your heart’s content. The Main Screen will now show the copy:



Title	State	Step	Timeout
Psyc209 Video Display 2	Ready	NA	NA
Psyc209 Video Experiment	Ready	NA	NA

Double-Click on any experiment to edit

Right-Click for other options.

Psyc209 Video Experiment

Step	Label	Type	Seq Number
5	Display Intro2A Bit...	eventDispl...	124
6	Wait for space key	keyinput	0
7	Display Intro3 Bitmap	eventDispl...	124
8	Wait for space key	keyinput	0
9	Display Intro4 Bitmap	eventDispl...	124
10	Wait for space key	keyinput	0
11	Label	label	19433760
12	Show Shapes	eventDispl...	0
13	Video	eventVideo	20859508
14	Wait for 1-8 Key Pr...	keyinput	2015353468
15	Branch	branchevent	0
16	Display End Text I...	eventDispl...	2015353468
17	Wait for space key	keyinput	20859508

This version will expire in 95 days.

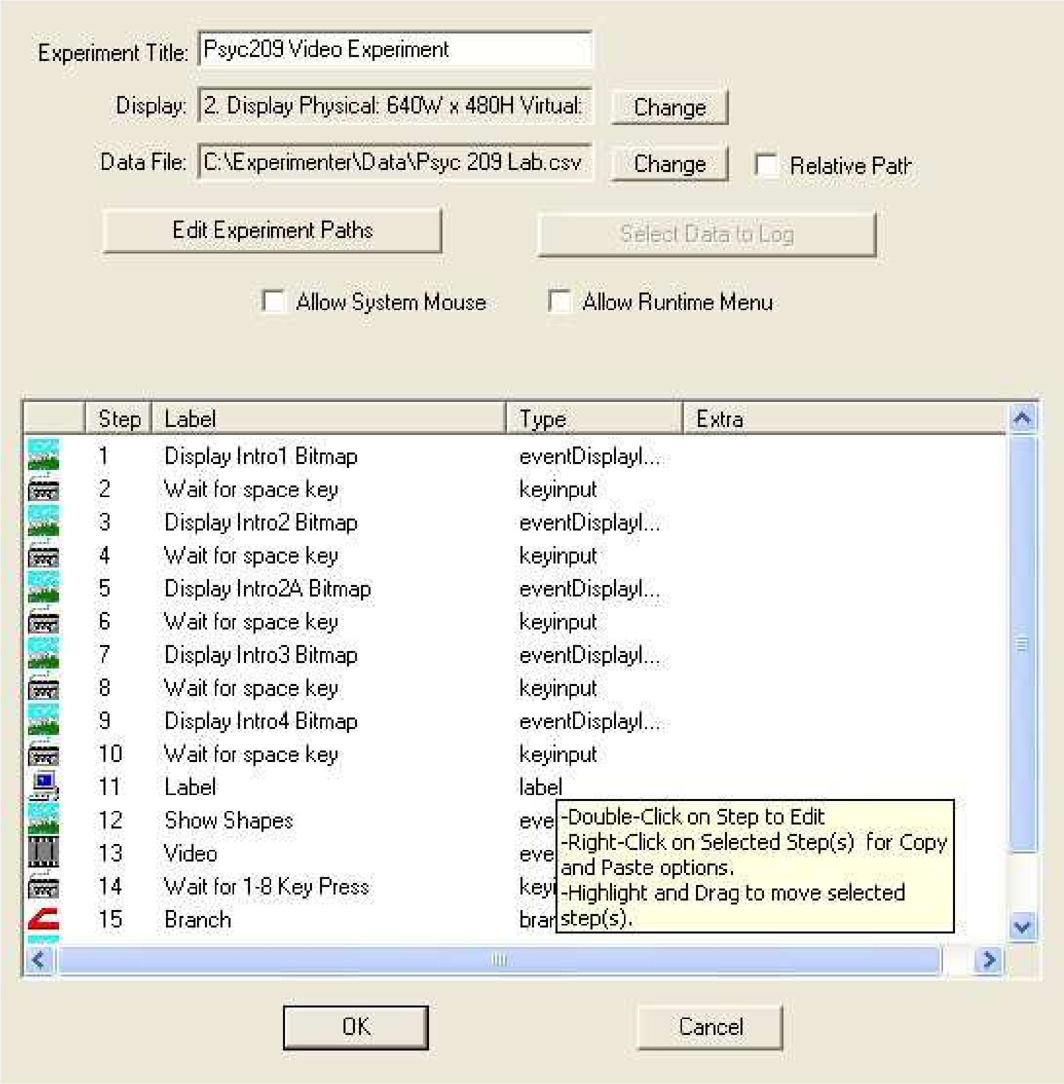
You may also make multiple copies of any and all experiments by using Shift-Click and Ctrl-Click in the left pane to select the experiment(s) to be copied and then following the procedure just outlined to make copies. Whichever experiment is selected in the left pane will be shown in the right pane.

Editing an Experiment

To edit an existing experiment simply Double-Click in the left pane on the experiment you wish to edit. To create and edit a new experiment Left-Click on the ‘New Experiment’ button. For this example we Double-Click on the first experiment to get to the Experiment Editor.

The Experiment Editor

This is how the Experiment Editor appears. This has an experiment already set up::



Experiment Title: Psyc209 Video Experiment

Display: 2. Display Physical: 640W x 480H Virtual:

Data File: C:\Experimenter\Data\Psyc 209 Lab.csv

☐ Relative Path

☐ Allow System Mouse ☐ Allow Runtime Menu

Step	Label	Type	Extra
1	Display Intro1 Bitmap	eventDisplayI...	
2	Wait for space key	keyinput	
3	Display Intro2 Bitmap	eventDisplayI...	
4	Wait for space key	keyinput	
5	Display Intro2A Bitmap	eventDisplayI...	
6	Wait for space key	keyinput	
7	Display Intro3 Bitmap	eventDisplayI...	
8	Wait for space key	keyinput	
9	Display Intro4 Bitmap	eventDisplayI...	
10	Wait for space key	keyinput	
11	Label	label	
12	Show Shapes	eve	
13	Video	eve	
14	Wait for 1-8 Key Press	keyi	
15	Branch	brar	

-Double-Click on Step to Edit
-Right-Click on Selected Step(s) for Copy and Paste options.
-Highlight and Drag to move selected step(s).

These are the items in the Experiment Editor:

- **Experiment Title –** this is the user-defined name for the experiment. This can be edited in place.
- **Display** - refers to the output display for all steps which have a visual display component. Clicking the ‘Change’ button brings up the Display Editor dialog.
- **Data File –** this is the destination file name for all data output when the experiment is run. Clicking the ‘Change’ button brings up a file dialog to set the data file name.

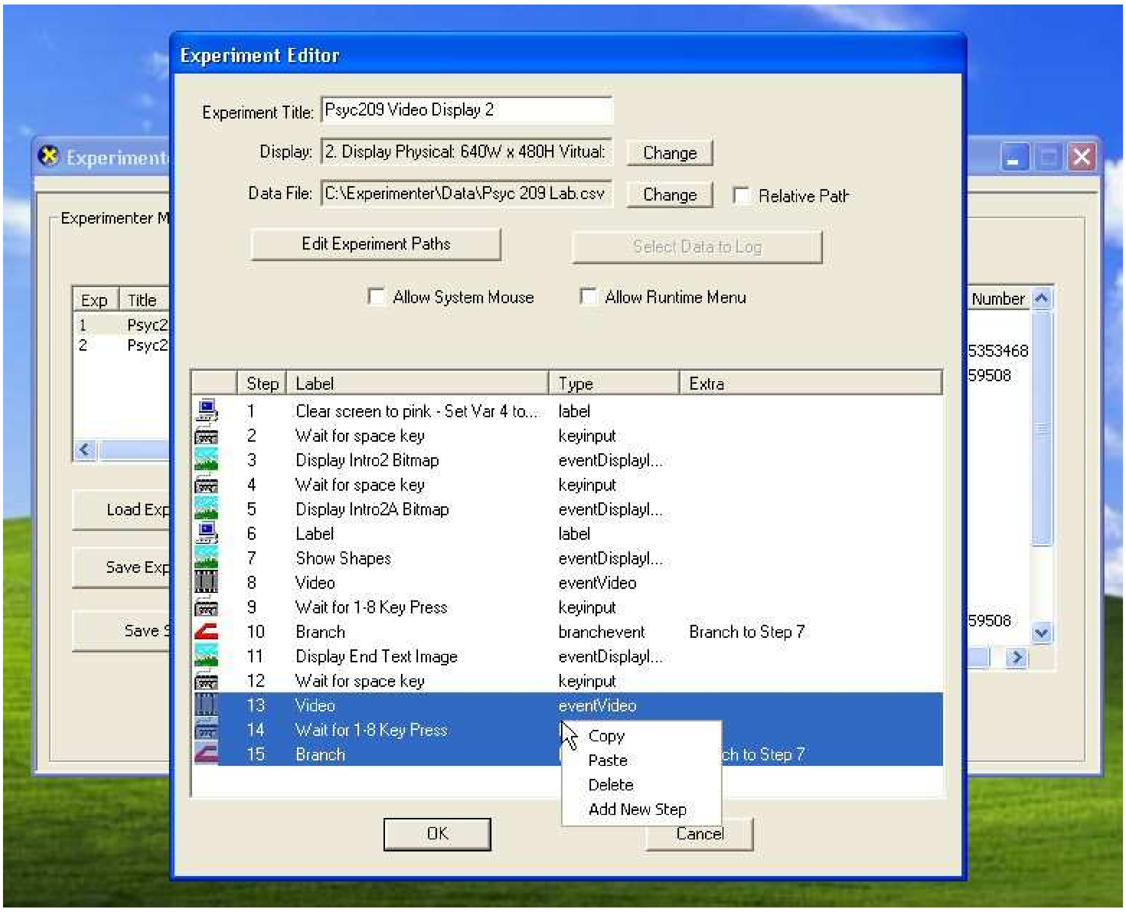
- **Edit Experiment Paths –** Opens the Editor for all default paths such as ‘Images’, ‘Data’, etc.

- **Allow System Mouse –** When the Experiment is running this allows the mouse to appear if checked. For a single display system deselecting this hides the mouse to prevent it ‘bleeding through’.

- **Allow Runtime Menu –** For multiple display systems you can monitor the experiment’s progress in the Runtime Menu by selecting this.

The large pane shows the steps in the current experiment. These are in the order in which they will be executed. The step number, user-defined label specified in the Step Editor and the Step Type are listed for each step. The ‘Extra’ column shows extra information such as the Step number that a Branch Step will branch to.

To copy steps, select the steps to copy with mouse Shift-Click or Ctrl-Click and then Right-Click to bring up the sub menu and select ‘Copy’ as shown below. Now Right-Click again and select ‘Paste’. A copy of the steps which you just copied and pasted will now be appended at the bottom of the steps list as shown (highlighted for illustration):



Experiment Title: Psyc209 Video Display 2

Display: 2. Display Physical: 640W x 480H Virtual:

Data File: C:\Experimenter\Data\Psyc 209 Lab.csv

☐ Relative Path

☐ Allow System Mouse

☐ Allow Runtime Menu

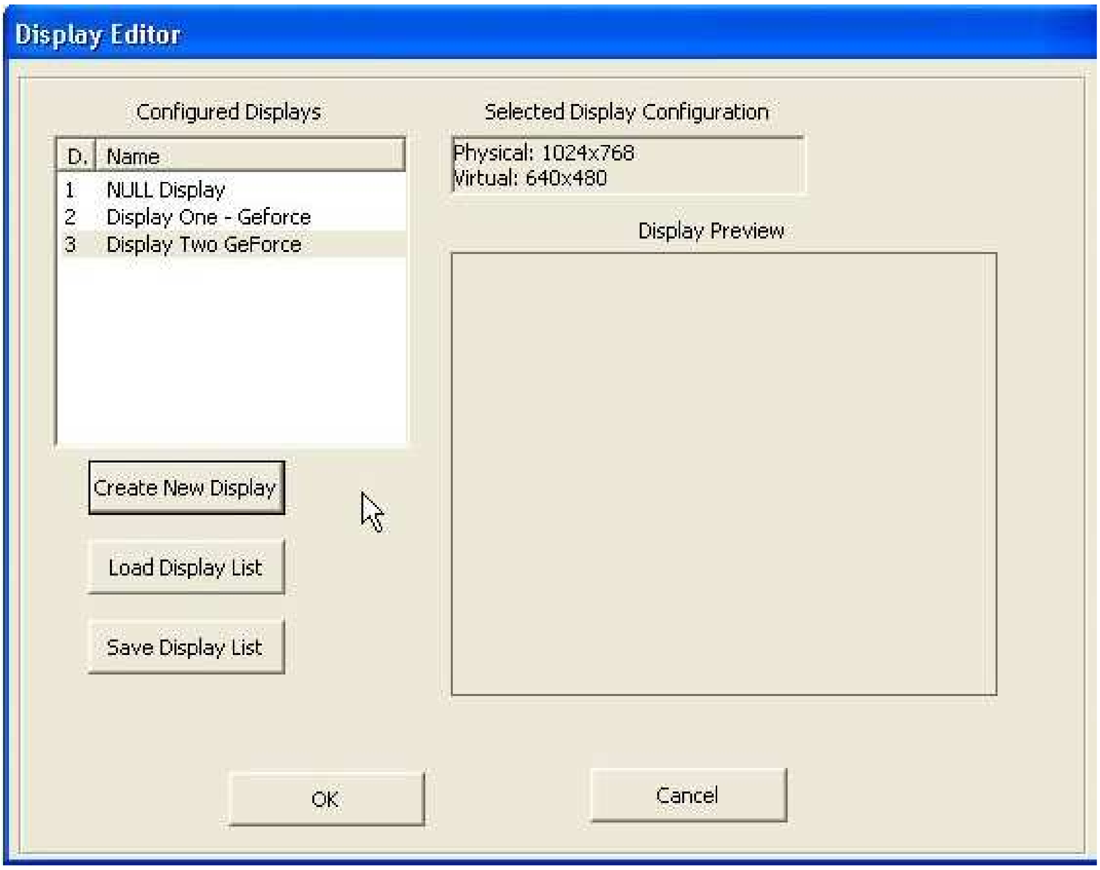
Step	Label	Type	Extra
1	Clear screen to pink - Set Var 4 to...	label	
2	Wait for space key	keyinput	
3	Display Intro2 Bitmap	eventDisplayI...	
4	Wait for space key	keyinput	
5	Display Intro2A Bitmap	eventDisplayI...	
6	Label	label	
7	Show Shapes	eventDisplayI...	
8	Video	eventVideo	
9	Wait for 1-8 Key Press	keyinput	
10	Branch	branchevent	Branch to Step 7
11	Display End Text Image	eventDisplayI...	
12	Wait for space key	keyinput	
13	Video	eventVideo	
14	Wait for 1-8 Key Press		
15	Branch		ch to Step 7

Steps can be moved by highlighting and then dragging the highlighted step(s) to the step under which you want to move the step(s). To move a step(s) to be the new first step(s), drag the highlighted items to the first step. Then highlight and drag the original first step down to its new proper position. Selection made be made in any combination using Shift Left-Click and CTRL Left-Click operations.

Display Configuration and Selection

Experimenter allows the concurrent use of as many display adapters as can be physically installed on one computer. Much of the early testing and development was done using a system with one dual-head AGP display adapter and 4 PCI display adapters for a total of six displays available (all six were actually used). The recent proliferation of USB 2.0 display adapters increases the number of display adapters possible, particularly for lower resolution display requirements.

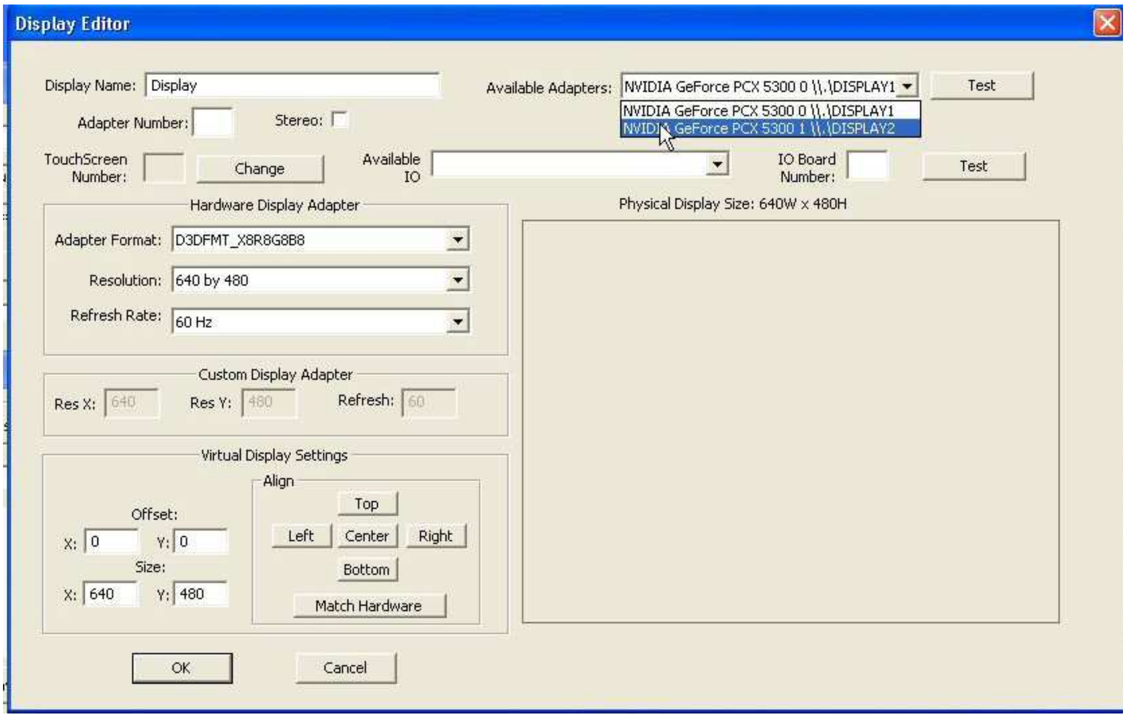
In the **Experiment Editor,** the **Display** box shows the currently selected display. Clicking on the **Change** button to the immediate right of that brings up the following **Display Selection Editor** window:



Physical: 1024x768
Virtual: 640x480

This menu allows the selection of the Display to use for this experiment configuration. When creating a new experiment there will only be the first entry, ‘NULL Display’. For the sake of illustration, two additional Displays have been configured. A virtually unlimited number of display configurations may be added. As in other *Experimenter* menus, any combination of configured Displays may be selected, copied, pasted or deleted at any time using the normal Windows Shift-Drag and Ctrl-Click key combinations.

There are two ways to activate the **Display Configuration Editor**: Double-Clicking on an existing Display in the **Configured Display** box and by selecting the **Create New Display** button. Performing these actions brings up the following:



Display Name: Display

Available Adapters: NVIDIA GeForce PCX 5300 0 \\.\DISPLAY1

NVIDIA GeForce PCX 5300 0 \\.\DISPLAY1
NVIDIA GeForce PCX 5300 1 \\.\DISPLAY2

Adapter Number:

Stereo: ☐

TouchScreen Number:

Available IO:

IO Board Number:

Physical Display Size: 640W x 480H

Adapter Format: D3DFMT_X8R8G8B8

Resolution: 640 by 480

Refresh Rate: 60 Hz

Res X: 640 Res Y: 480 Refresh: 60

Offset: X: 0 Y: 0

Size: X: 640 Y: 480

This menu provides the following options:

- **Available Adapters –** This is a list of the display adapters ***that are physically present*** on the system. No more or less.
- **Test –** selecting this will display a test image on the currently selected adapter
- **Display Name –** This can be any descriptive name.
- **Hardware Display Adapter Adapter Format –** This is a drop-down list of the colour formats supported by the currently selected adapter.
- **Hardware Display Adapter Resolution –** This shows all screen resolution that this adapter supports. It will not show any unsupported resolutions.
- **Hardware Display Adapter Refresh Rate –** This will show any refresh rates actually allowed for the current adapter at the currently selected **Resolution**.
- **Custom Display Adapter X/Y/Refresh –** These settings allow a user-defined display adapter. This is included for development of experiments intended to be run on a different computer.

- **Virtual Display Settings Offset/Size X/Y –** These settings allow a user to have virtual displays within the actual physical display contexts. The idea here is to allow separate experiments to run concurrently using different portions of the same physical display adapter.

- **Virtual Display Settings Align Options –** These are helpers to allow rapid placement of **Virtual Displays**.

- **Match Hardware –** this forces the **Virtual Display** setting to match the **Physical Display** selections.

The Sequence Editor and Sequence Lists

The concept of the **Sequence List** is key to *Experimenter* and is frequently mentioned throughout this document. A **Sequence List** is simply a list of similar items such as a list of text strings or a list of numbers or a list of RGB colour values. These are used primarily to repeatedly cycle through a block of Steps, substituting different values from the **Sequence List** for variables in the Step. One example would be to display different images each time a particular Step is executed by having a Sequence List of filenames.

Sequence Lists may be defined in one of two ways: as a plain-text Microsoft Excel .CSV file and as an automatically-generated Sequence List. CSV files may consist of multiple columns where each column is a separate Sequence List. Columns are comma delimited and rows are delimited by the ‘newline’ character. The only restriction at present is that each column must have the same number of items. Adjacent columns may consist of any type of data, although the type needs to be the same for the entire column.

Columns may be further broken down into categories. An example would be the case where 20 images are to be presented. A column is created in any text editor or in Excel with the first 10 items being filenames for a particular category of image, say images of birds. The next ten items are pictures of dogs. We configure the list to be presented in random order and want to log data for which category was presented for matching to a subsequent key press. We simply set the number of categories to ‘2’ and *Experimenter* automatically tracks which category is presented each time even though the ordering has been convoluted.

If two Sequence Lists are used and the random number seed for each list is the same then there is a guarantee that the lists will be randomized in the same order. In the example just given we may have a second column of numbers representing the expected key press for a subsequent **Keyboard Input** step. If the seed is the same, ***even across multiple steps***, then the rows, although presented in random order, will always retain their relative ordering. This is true whether different columns in one CSV file are used, different CSV files or any combination.

For RGB Colour Sequence Lists, each item is assumed to occupy 3 consecutive columns in the same CSV file representing the Red, Green and Blue values respectively. Values should be 0-255. Higher values will be modulus 255.

In the **Step Editor** screens for the various step types certain variables will have a yellow or green background. This signifies that the variable is 'Sequenceable'. The following snippet from the Display Image Step Editor illustrates this:

Display Image Step Editor

Source Image Attributes

File: t\NEW_PROJ\IMAGES\BUTTONS.BMP

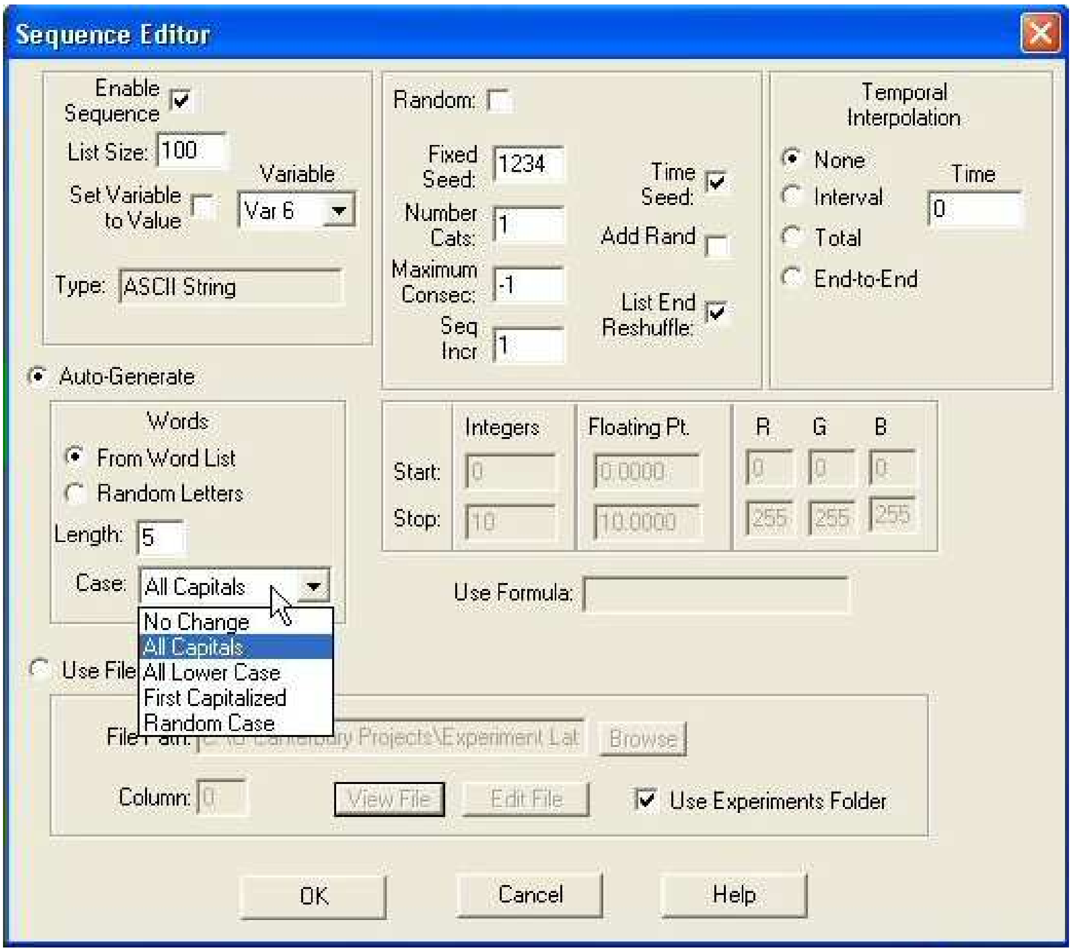
Relative Path ☐ Width: 640 Height: 480

Output Image Display Attributes

Resize ☐ Width: 640 Height: 480 ☑ Maintain Aspect

Note that the **File, Width** and **Height** boxes all have yellow background. Right-Clicking on any of these Sequenceable variables gives the option to 'Edit Sequence List'. If that option is selected then the Sequence Editor for that variable is displayed. A **yellow** background indicates that the actual value entered in the box will always be used. If we enable the sequence in the Sequence Editor, then this will change to a **green** background in which case the Sequence List is used and the value in the box is ignored.

When **Edit Sequence List** is selected for a variable, the **Sequence Editor** screen pops up:



Enable Sequence ☑
List Size: 100
Set Variable to Value ☐
Variable: Var 6
Type: ASCII String
Random: ☐
Fixed Seed: 1234
Number Cats: 1
Maximum Consec: -1
Seq Incr: 1
Time Seed: ☑
Add Rand ☐
List End Reshuffle: ☑
● None
○ Interval
○ Total
○ End-to-End
Time: 0
● Auto-Generate
● From Word List
○ Random Letters
Length: 5
Case: All Capitals
Integers Start: 0, Stop: 10
Floating Pt. Start: 0.0000, Stop: 10.0000
R G B Start: 0 0 0, Stop: 255 255 255
Use Formula:
○ Use File
Column: 0
☑ Use Experiments Folder

This is a description of configuration options in the **Sequence Editor**:

- **Enable Sequence –** if this is checked then the Sequence List will be used for the Step configuration option that we selected to get to this screen. If not checked then the Sequence List will NOT be used.
- **List Size –** this option is used only when the Auto-Generate radio button is ticked. This is the number of items created in an auto-generated list.
- **Set Variable to Value –** This item is only meaningful for key and numeric lists. When the parent step is executed it will set the selected Global Variable to the current Sequence List value.
- **Type –** Read-Only – this shows the type of the variable, i.e. Integer, RGB, etc.
- **Random –** if checked then the list will be presented in pseudo-random order.
- **Fixed Seed –** This is a fixed number that determines the ordering of the randomized list. All lists using this seed, assuming they are the same length, will be randomized in the same order.
- **Time Seed –** if checked then the *Experimenter* program start time value is added to the **Fixed Seed** value. This allows different lists to be randomized in

the same order as each other, but that order will be different each time the program is run.

- **Number Cats –** Number of categories to divide the list into.
- **Maximum Consec –** maximum number of consecutive items of the same category to allow.
- **Seq Incr –** meaningful only if temporal interpolation is not being used. This is the number of list items to increment each time the step is executed. Normally set to 1.
- **List End Reshuffle –** the normal behaviour for a Sequence List is to ‘wrap’ when the list is exhausted. This means that if a step has a Sequenced variable with a list size of 10 items and the step is executed 11 times then it will start at the beginning of the list on the 11th iteration. If **List End Reshuffle** is checked then the list will be randomized again when exhausted. This does not affect correlated lists – they will always reordered the same.
- **Temporal Interpolation None –** checked - Sequence Lists are always executed sequentially. **Every** time the step is executed the Sequence List counter is incremented.
- **Temporal Interpolation Interval –** checked **-** the Sequence List counter is incremented only is the time specified in the **Time** box has elapsed since the last time this step incremented.
- **Temporal Interpolation Total –** a total time is specified in the **Time** box. This is the total time that the list should be completely sequenced. For example, the Sequence List has 100 elements which should be completely exhausted in 2000 milliseconds. Each time the step is executed the current item in the list to be used is calculated from the first time the step was executed. This means that if 1001 ms has elapsed then the item at position 50 would be used. If next time 1008 ms has elapsed item number 50 would be used again. If the next time 1017 ms has elapsed then we use the item at position 51, because that is the nearest expected item. This is useful for smooth movement of bitmap images across the screen.
- **Temporal Interpolation End-to-End –** this is only valid for auto-generated numerical Sequence Lists. It works the same as the **Total** except the actual value to be used is calculated instead of the position in the list. This is based on **Start/Stop** values specified in the **Auto-generate** section values and the value in **Time**. If the **Start** value is 12, the **Stop** value is 24, and the **Time** value is 1000 then the value used would be calculated as 12 + elapsed * (24-12)/1000)
- **Auto-Generate From Word List –** create a word list using the built-in 10,000+ most common English words list. This uses the value in the **List Size** box to determine number of words to grab.
- **Auto-Generate Random Letters –** create a word list using random letters. This uses the value in the **List Size** box to determine number of ‘words’ to generate.

- **Auto-Generate Length –** length of word taken from word list or created from random letters..

- **Auto-Generate Case –** letter case to use in word lists as shown in the example above..

- **Auto-Generate Integer/Floating Pt. Start/Stop –** for numerical values this will be the first and last values for the list. The list will be the length of the value in the **List Size** box. A linear interpolation is used to determine the value of each item in the list. List may be low-high or high-low.

- **Auto-Generate R/G/B Start/Stop –** a list of RGB values is created using the value in the **List Size** box for the size of the list and performs a linear interpolation to create items. If values larger than 255 are used in the Start or Stop boxes then the value will be the calculated value modulus 255. To create a complete grey-scale list use a list size of 256 with RGB Start values of 0, 0, 0 and Stop values of 255,255,255.

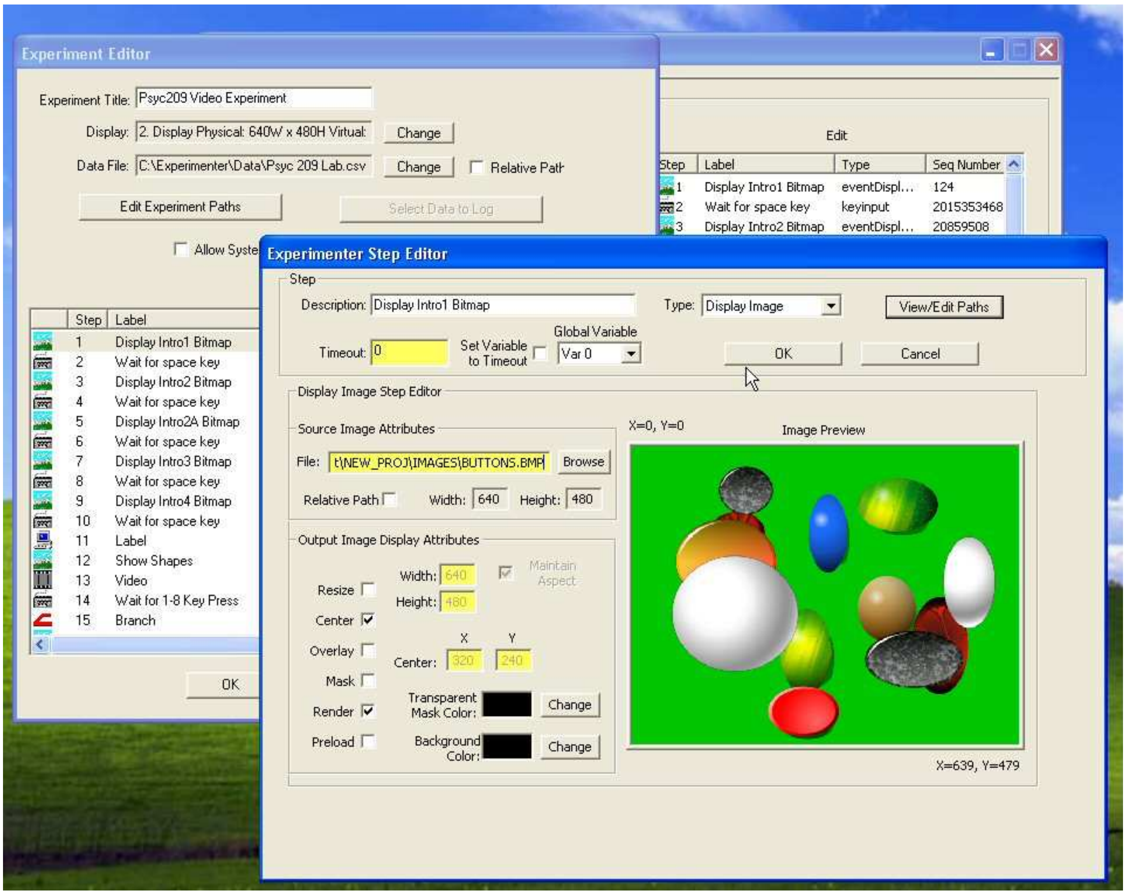
- **Auto-Generate Case –** letter case to use in word lists as shown in the example above..

Editing Steps

Getting to the Step Editor is similar to getting to the Experiment Editor from the Main Screen. To edit an existing step, Double-Click on step the step to edit. To add a new step, just Right-Click and select ‘Add New Step’ from the drop-down menu.

The Step Editor

The Step Editor looks similar to the following:



Experiment Title: Psyc209 Video Experiment
Display: 2. Display Physical: 640W x 480H Virtual:
Data File: C:\Experimenter\Data\Psyc 209 Lab.csv
☐ Relative Path

Step	Label
1	Display Intro1 Bitmap
2	Wait for space key
3	Display Intro2 Bitmap
4	Wait for space key
5	Display Intro2A Bitmap
6	Wait for space key
7	Display Intro3 Bitmap
8	Wait for space key
9	Display Intro4 Bitmap
10	Wait for space key
11	Label
12	Show Shapes
13	Video
14	Wait for 1-8 Key Press
15	Branch

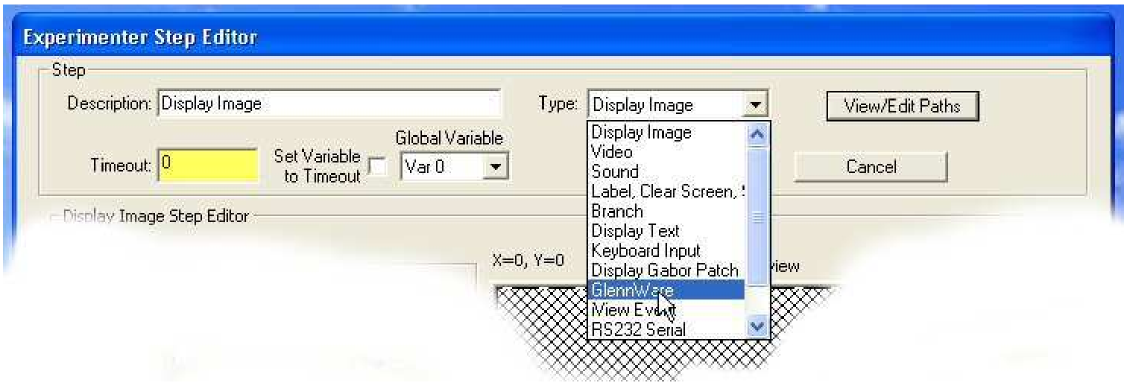
Step	Label	Type	Seq Number
1	Display Intro1 Bitmap	eventDispl...	124
2	Wait for space key	keyinput	2015353468
3	Display Intro2 Bitmap	eventDispl...	20859508

Description: Display Intro1 Bitmap
Type: Display Image
Timeout: 0
Set Variable to Timeout ☐
Global Variable: Var 0

File: t\NEW_PROJ\IMAGES\BUTTONS.BMP
Relative Path ☐
Width: 640
Height: 480

Resize ☐
Center ☑
Overlay ☐
Mask ☐
Render ☑
Preload ☐
Width: 640
Height: 480
☑ Maintain Aspect
Center: X 320 Y 240
Transparent Mask Color: ■
Background Color: ■

The Step Editor Dialog consists of two sections. The top section is the same for every type of step. The bottom section is specific to the type of step. In the example above the bottom section is specific to the 'Display Image' Step. Clicking in the top section's **Type** window would display the following. Note that this installation of *Experimenter* has several Step Plugins in addition to the basic Step types. Any Step type shown in the Type pull-down menu may be selected. It is also possible to change the type of an existing Step here.



Description: Display Image
Type: Display Image
Timeout: 0
Global Variable: Var 0

Editing the **Description** box will change the label of this particular step in the Experiment Editor. This can be as descriptive as needed.

The **Timeout** box is the time in milliseconds that this step will take to execute when the experiment is running. If this step is a Display Image step with a timeout value of 1000, then when the experiment is running and it reaches this step the image will be displayed for 1000 milliseconds (1 second) at which time the next step will execute. If the timeout were 0 milliseconds then the image would be displayed and the next step would immediately execute.

Timeouts are not used for every type of step. When a Branch Step is executed it immediately branches or else moves on to the next step regardless of the timeout value. When a Key Input Step is executed the step ends when an allowed key is pressed or at the timeout, whichever comes first.

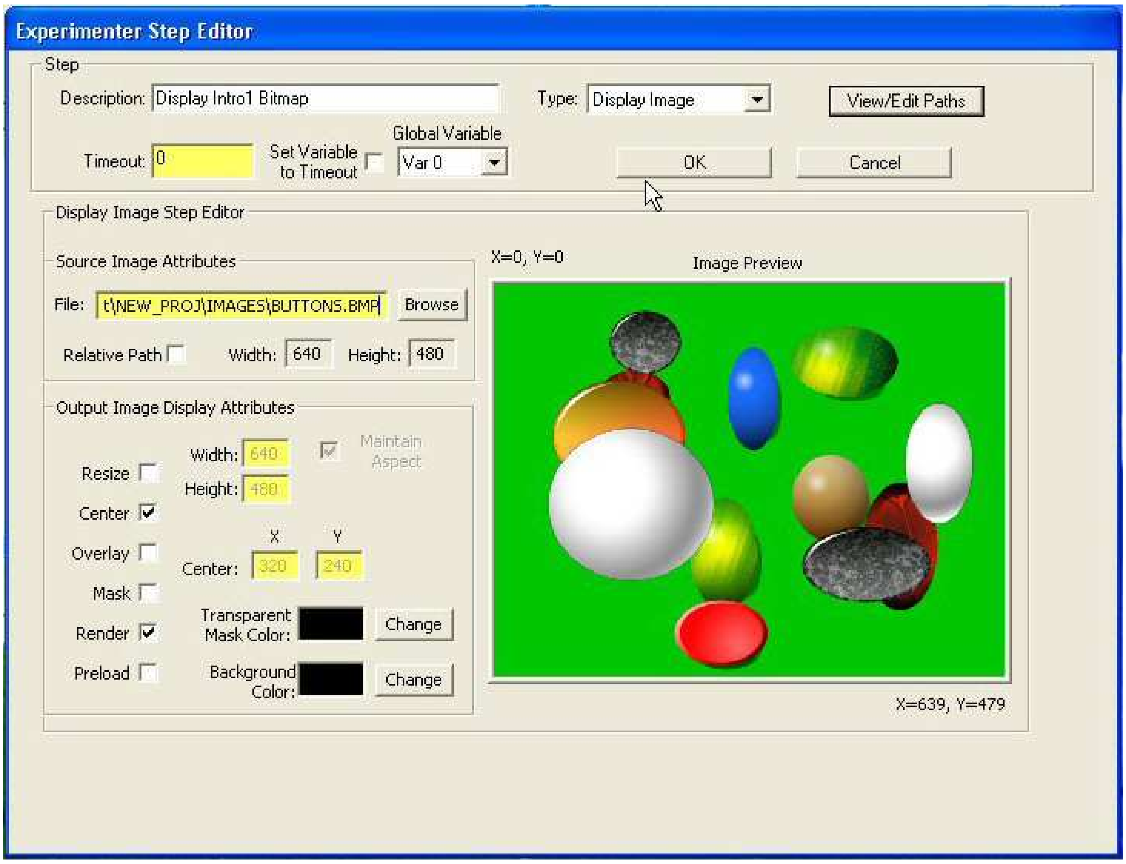
Note that the Timeout box has a yellow background. This means that it is a ‘Sequenceable’ value. This is to allow different timeouts each time this particular Step is executed.

Checking the **Set Variable to Timeout** box will set the selected Global Variable to the value used when this step is executed. This is useful when the timeout value uses a Sequence and a later Branch step depends on the timeout value.

The **View/Edit Paths** button is here as a convenience and duplicates the action of the **View/Edit Paths** button in the Experiment Editor.

Display Image Step Editor

This is a view of the Display Image Step Editor:



Description: Display Intro1 Bitmap
Type: Display Image
Timeout: 0
Set Variable to Timeout ☐
Global Variable: Var 0
File: t\NEW_PROJ\IMAGES\BUTTONS.BMP
Relative Path ☐ Width: 640 Height: 480
Resize ☐
Center ☑
Overlay ☐
Mask ☐
Render ☑
Preload ☐
Width: 640
Height: 480
☑ Maintain Aspect
Center: X 320 Y 240
X=0, Y=0
X=639, Y=479

This is a description of the various fields and option for this step:

- **File Path** – This is the path of the image file to be displayed when this step is executed. Clicking on the ‘Browse’ button opens a file browser displaying all supported image file formats. Once an image has been selected it is displayed in the preview pane. The actual dimensions of the image are shown in the ‘Width’ and ‘Height’ boxes under the file path. This field is Sequenceable so that different filenames can be used each time the step is executed.

- **Relative Path** – If checked then the filename is parsed from the full path and the default image path will be prepended to the filename. This is to enable portability between systems and individual experiments, letting the images reside in one folder on one computer and a different folder on another computer.

- **Width –** This is read-only and is the actual width of the source image when not using a sequence.

- **Height –** This is read-only and is the actual height of the source image when not using a sequence.

- **Resize –** Checking this will force the image to be rendered in the specified size.

- **New Width –** Width to resize image to

- **New Height –** Height to resize image to.

- **Center –** Checking this will cause the image to be displayed in the centre of the screen

- **Overlay –** If checked then the existing screen will be retained and the image drawn over it, otherwise the existing screen will be cleared first.

- **Mask –**Checking this will cause any portion of the image which is the same colour as the 'Transparent Colour' to be transparent.

- **Render** - If checked then the image will be rendered immediately upon step execution, otherwise it is simply buffered in an off-screen buffer. This is useful when consecutive Display Image Steps must be displayed simultaneously. All but the last of several Display Image steps would have the 'Render' checkbox unchecked and 'Retain Existing Screen' checked. The last 'Display Image Step' in the series would have the 'Render' checkbox checked, causing it and the preceding steps to display in the same screen refresh cycle. Text and video may also be layered in similar fashion.

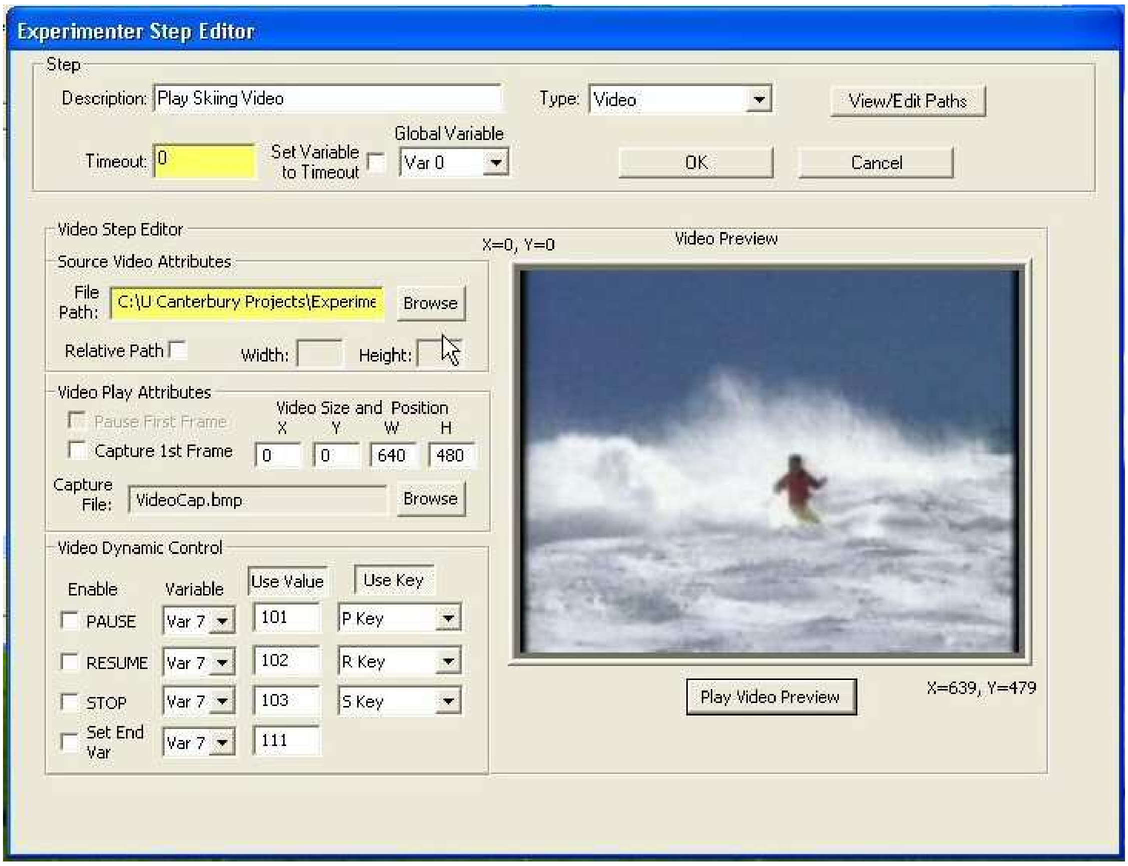
- **Center X & Y –** These specify the screen location of the centre of the image. Nice for displaying moving cursors or displaying stimuli in varying screen locations.

- **Transparent Color –** The transparent color if the 'Mask checkbox is checked. This can be used for creating Sprites and Masks. Clicking the 'Change' button brings up the standard color selection palette:

If for example the colour red (RGB 255,0,0) is selected here then any pixels in the image with the same value would be transparent, showing whatever pixel colour was already there.

Video Step Editor

The Video Step Editor allows the user to load video images and preview them. When Video Steps are executed they will run until either the video clip finishes or a timeout takes place. Most times the Timeout value will be set to a large value as in the example so that the video clip will always finish.



Description: Play Skiing Video
Type: Video
Timeout: 0
Set Variable to Timeout ☐
Global Variable: Var 0
File Path: C:\U Canterbury Projects\Experime
Relative Path ☐
Width:
Height:
☐ Pause First Frame
☐ Capture 1st Frame
X: 0 Y: 0 W: 640 H: 480
Capture File: VideoCap.bmp

Enable	Variable	Use Value	Use Key
☐ PAUSE	Var 7	101	P Key
☐ RESUME	Var 7	102	R Key
☐ STOP	Var 7	103	S Key
☐ Set End Var	Var 7	111	

The following configuration options for the Video Step are available:

- **Browse –** Click this button to bring up a file dialog. Most common video file formats such as AVI, WMV, etc. are recognized by *Experimenter*.

- **Video Size and Position –** Video may be resized and played back anywhere on the display for this experiment. Multiple videos can be played at the same time by using more than one Video Step.

- **Play Video Preview –** This will play the loaded video in the preview window.

- **Pause/Resume/Stop –** Any combination of these options may be used to control playback. For example, checking Stop with the default Var 7 = ‘S Key’ means that once the video is playing, the callback function will monitor for user keypresses and if the ‘S’ key is pressed the video will immediately stop playing.

- **Capture First Frame –** This is a simple utility that when checked will automatically save the first frame of a video clip as a BMP with the name, CAPxxxxxx.BMP where ‘xxxxxx’ represents a random number. This is useful in cases where the *Experimenter* may want to freeze the first frame of a video clip for a fixed period of time or until a key is pressed and then seamlessly play the rest of the video. This would be done by:

    1. Select ‘Add New Step’ in the Experiment Editor.

    2. Choose ‘Video’ in the ‘Type’ dropdown list in the Step Editor.

    3. Select ‘Browse’ and select the video file.

    4. Check the ‘Capture First Frame’ check box.

    5. Save the experiment and then run it once. You can terminate the experiment anytime after the first frame of the video has been displayed.

    6. Create a new Display Image step in your experiment. Edit the step and for the filename select the most recent file in your *Experimenter* folder with the CAPxxxxxx.BMP name.

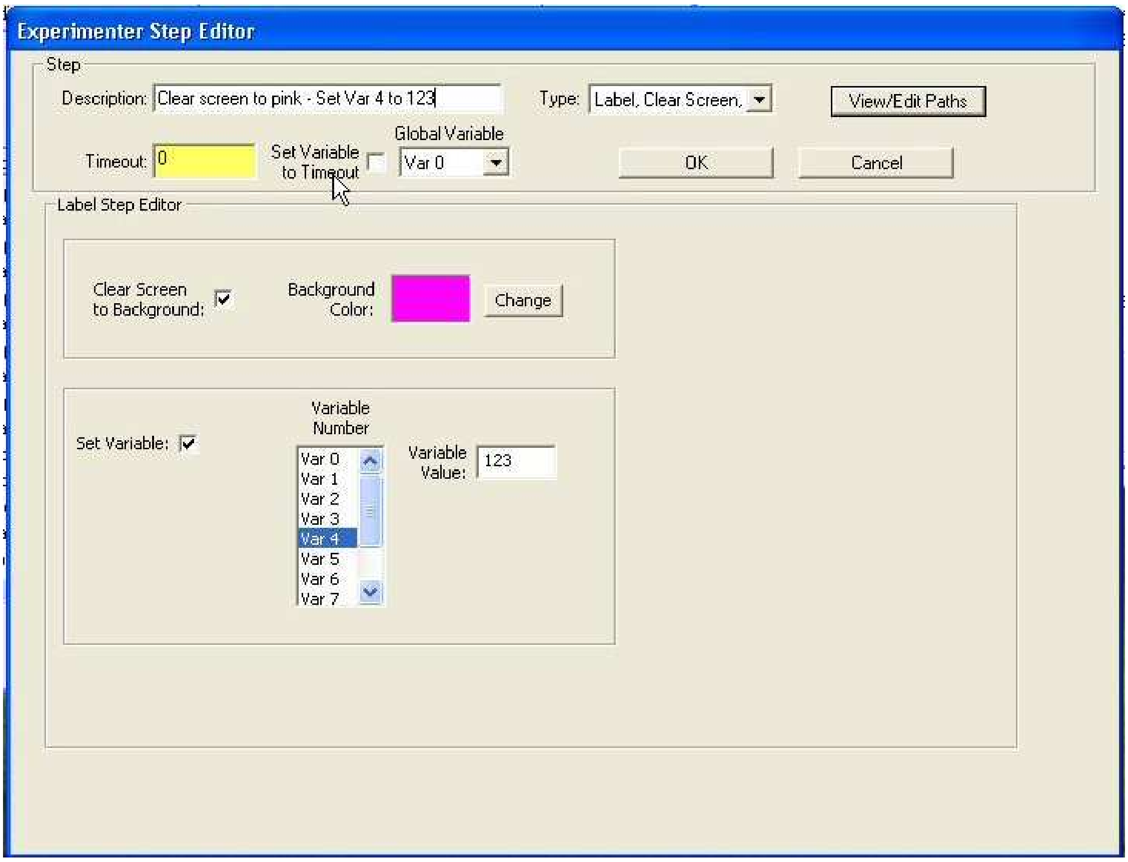
    7. Place the new Display Image step immediately before the Video step and set the Timeout value for the Display Image step to the delay time desired. When the experiment is next run there will be a flawless transition from the ‘paused’ first frame to the real vide.

    8. If a keypress pause is desired, simply insert a Keyboard or Mouse Input step between the Display Image and Video steps and set the Timeout value in the Display Image step to 0. When the experiment is run the first frame will be displayed until the key is pressed and then the video will continue playing.

Label Step Editor

The Label Step provides some miscellaneous features. The most common use for the Label Step is to clear the screen when the experiment is running.

Example Label Step



Description: Clear screen to pink - Set Var 4 to 123

Type: Label, Clear Screen,

Timeout: 0

Set Variable to Timeout ☐

Global Variable: Var 0

Clear Screen to Background: ☑

Background Color:

Set Variable: ☑

Variable Number: Var 4

Variable Value: 123

The following **Label Step** items are configurable:

- **Clear Screen to Background –** If this is checked then when the experiment is run and this step is executed the screen will be cleared to the colour displayed in the 'Background Colour' box.

- **Background Colour –** This is changed by clicking on the 'Change' button and then choosing a colour from the Colour Selection Palette.

- **Set Variable –** If this box is checked then when this step is executed, the Global Variable selected in the 'Variable Number' list will be set to the value in the 'Variable Value' edit box.

- **Variable Number –** Select box for Global Variables.

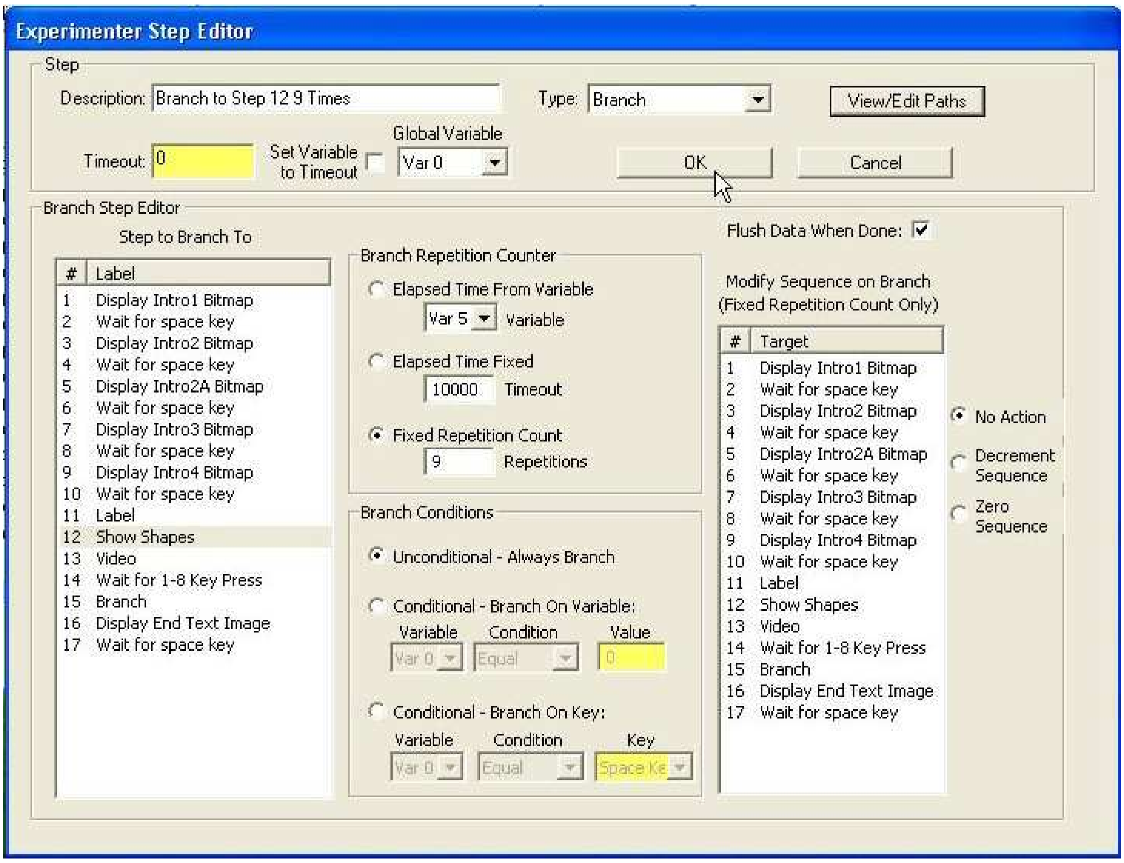
- **Variable Value –** Edit box for a number to apply to the selected Global Variable.

To simply clear the screen during an experiment, perhaps after a Display Image or Video step, simply insert a Label Step with a Timeout value of 0 milliseconds and with the ‘Clear Screen to Background’ checkbox checked. Alternatively you may wish to clear the screen and wait for a specified time in which case simply set the Timeout value for the required delay.

Note the judicious use of the Description box. Whatever is entered here will be displayed in the Steps List in the Experiment Editor and on the Main Screen. This provides the opportunity to clearly document what each step is doing.

Branch Step Editor

The Branch Step is one of the most powerful features of *Experimenter*. It provides a great deal of flexibility in the flow of the experiment. While simple experiment programs execute in a ‘toe-to-heel’ fashion, *Experimenter* allows new paths to be taken based on dynamic inputs and changing conditions. An experiment could contain a Keyboard Input step which only allows the ‘1’, ‘2’, ‘3’ and ‘4’ keys and then based on which key was pressed a completely different sequence of steps could be executed. This is an example of the Branch Step:



Step

Description: Branch to Step 12 9 Times

Type: Branch

Timeout: 0

Set Variable to Timeout ☐

Global Variable: Var 0

Branch Step Editor

Step to Branch To

#	Label
1	Display Intro1 Bitmap
2	Wait for space key
3	Display Intro2 Bitmap
4	Wait for space key
5	Display Intro2A Bitmap
6	Wait for space key
7	Display Intro3 Bitmap
8	Wait for space key
9	Display Intro4 Bitmap
10	Wait for space key
11	Label
12	Show Shapes
13	Video
14	Wait for 1-8 Key Press
15	Branch
16	Display End Text Image
17	Wait for space key

Branch Repetition Counter

○ Elapsed Time From Variable
Var 5 Variable

○ Elapsed Time Fixed
10000 Timeout

● Fixed Repetition Count
9 Repetitions

Branch Conditions

● Unconditional - Always Branch

○ Conditional - Branch On Variable:
Variable: Var 0 Condition: Equal Value: 0

○ Conditional - Branch On Key:
Variable: Var 0 Condition: Equal Key: Space Ke

Flush Data When Done: ☑

Modify Sequence on Branch (Fixed Repetition Count Only)

#	Target
1	Display Intro1 Bitmap
2	Wait for space key
3	Display Intro2 Bitmap
4	Wait for space key
5	Display Intro2A Bitmap
6	Wait for space key
7	Display Intro3 Bitmap
8	Wait for space key
9	Display Intro4 Bitmap
10	Wait for space key
11	Label
12	Show Shapes
13	Video
14	Wait for 1-8 Key Press
15	Branch
16	Display End Text Image
17	Wait for space key

● No Action
○ Decrement Sequence
○ Zero Sequence

There are currently three types of Branch Steps. All three share the following configuration options:

- **Step to Branch To –** this is the step that will be executed next if the branch is taken; the ‘Target’ step. This window lists every step for this experiment. Any step may be selected, including this Branch Step (useful for empty time-out loops). *Experimenter* internally assigns a unique identifier to each step as it is created. The Branch step always branches to the step with that identifier.

  What this means is that if this Branch Step was configured to branch to the first step, ‘Start Text’ in the example above, then any number of further changes could be made to the experiment, but this Branch Step will always branch to the original ‘Start Text’ step. You can insert or delete steps (except for the target step), move this Branch step or the target step anywhere in the Step List ordering or copy these steps and this Branch step will still branch to the same step.

- **Elapsed Time From Variable–** if selected then the first time this step is reached, the current value of the selected Global Variable will serve as a timeout value (in milliseconds). The branch condition (if any) will be evaluated every time this step is executed until the timeout value has been reached. It then ‘falls through’ to the next step and resets.

- **Elapsed Time Fixed–** similar to the **Elapsed Time FromVariable,** except that the value used is taken from the **Timeout** box.

- **Fixed Repetition Count –** if selected then this is the number of times that this step will execute. If this number is ‘3’ then the 4th time this step is reached while the experiment is running it will ‘fall through’ to the next step in the Steps List and reset.

- **Step(s) to Modify Sequence on Branch Taken –** Later in this document we will be describing in detail *Experimenter*’s robust ‘Sequencer’. In brief, the ‘Sequencer’ allows the experiment designer to assign ‘Sequences’ to many of the configuration variables for the various steps. These ‘Sequences’ allow the variables to change value each time the step is executed. A Display Image Step could use a Sequence for the image file name so that every time that step is executed a different image is displayed.

  It may be that the researcher would like to have an image (or other variable) repeated when an incorrect key is pressed. By selecting the Display Image Step and checking the ‘Decrement Sequence’ radio button the sequence counter for that Display Image Step will be decremented to the previous value. All Sequence variables for all steps selected in this list will have the desired action applied if the branch is taken.

- **No Action –** This is the default. Selecting this means that no Sequence variables are modified if the branch is taken.

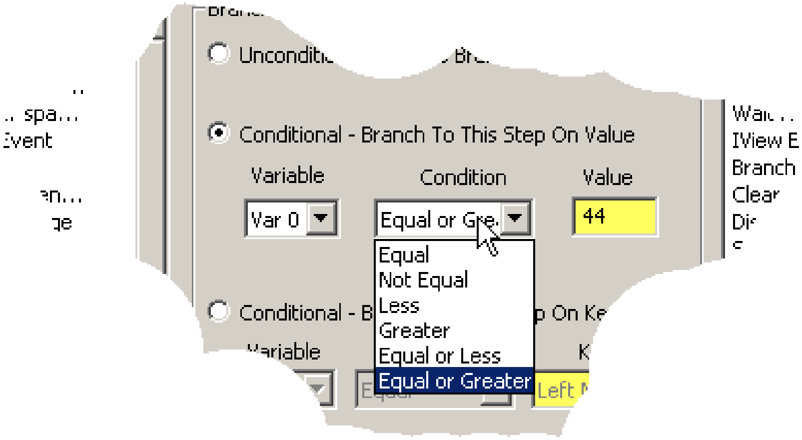
- **Decrement Sequence –** Decrements the Sequence Counters for all selected steps if the branch is taken.

- **Reset Step Sequences –** Resets the Sequence Counter to the beginning of the sequence for selected steps if the branch is taken.

Branch Conditions

- **Unconditional – Always Branch To This Step –** This selection always causes a branch to the selected step.

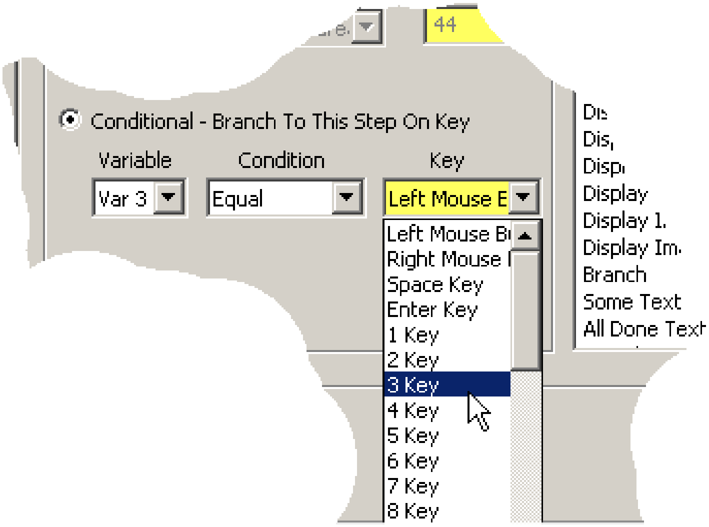
- **Conditional – Branch to This Step On Value –** This is a conditional branch comparing the Global Variable selected in the ‘Variable’ list to the number in the ‘Value’ edit box using the evaluation formula specified in the ‘Condition’ dropdown list.



● Conditional - Branch To This Step On Value

Variable: Var 0
Condition: Equal or Gre.
Value: 44

- **Conditional – Branch to This Step On Key –** This is a conditional branch comparing the Global Variable selected in the ‘Variable’ list to the ‘key’ value in the ‘Key’ dropdown list using the evaluation formula specified in the ‘Condition’ dropdown list. These key values are the result of a key press or mouse click in a previous ‘Keyboard or Mouse Input’ step.



● Conditional - Branch To This Step On Key

Variable: Var 3
Condition: Equal
Key: Left Mouse E

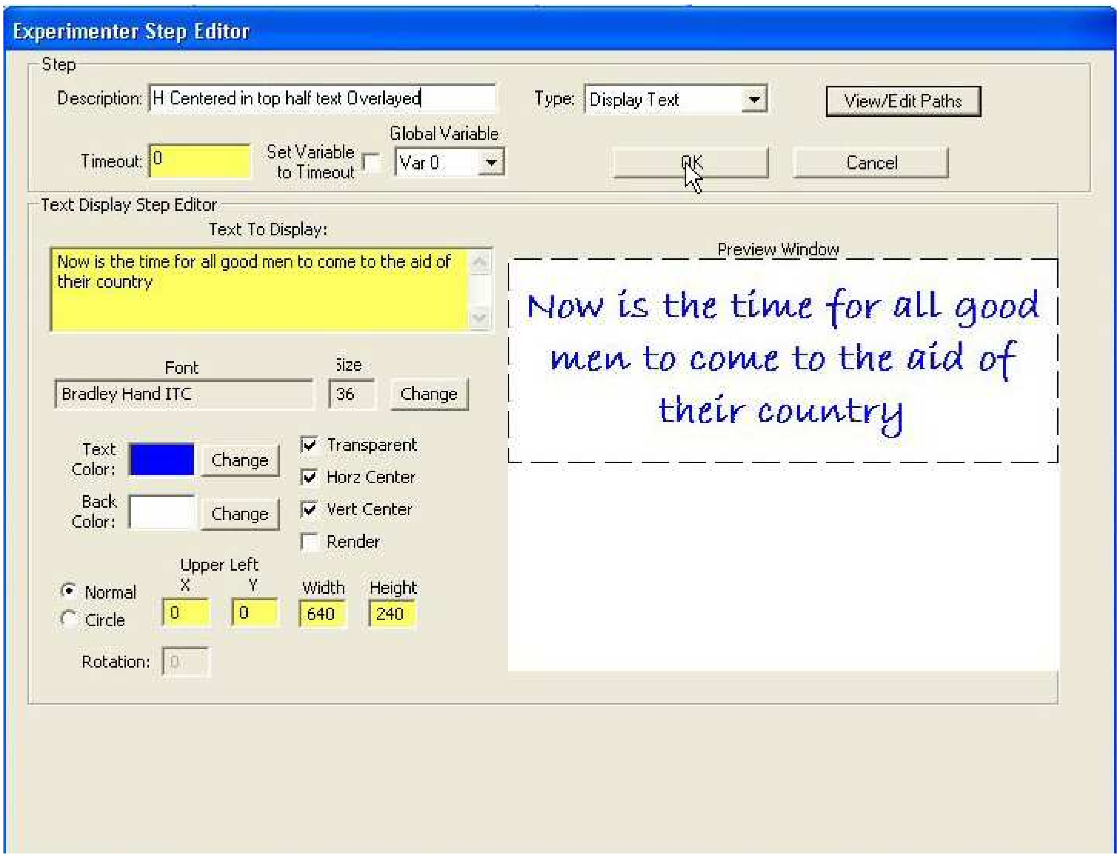
Additional Notes on the Branch Step

Branch Steps may be nested without any real limit. Whenever a Branch Step times out or executes for a fixed repetition count, it will reset to its initial state. If there is an inner Branch set to repeat 9 times and an outer branch set to repeat 4 times then the innermost group of steps would be executed a total of 50 times – 10 times for the inner branch group multiplied by 5 times for the outer branch. Some care needs to be exercised to prevent jumping outside of an outer Branch from an inner Branch.

Display Text Step

This step allows text strings to be displayed in arbitrary font styles, fore/background colours, font sizes and screen locations. In addition text may be displayed in a circle or at different orientations as well as the normal left-to-right horizontal layout. For normally laid-out text the experiment designer specifies a bounding rectangle for the text. *Experimenter* will automatically perform word wrap if necessary. Text may be centred horizontally, vertically or both.

The following example shows a Text Step that has been modified:



Description: H Centered in top half text Overlayed

Type: Display Text

Timeout: 0

Set Variable to Timeout ☐

Global Variable: Var 0

Text To Display: Now is the time for all good men to come to the aid of their country

Font: Bradley Hand ITC

Size: 36

☑ Transparent

☑ Horz Center

☑ Vert Center

☐ Render

● Normal

○ Circle

Upper Left X: 0 Y: 0

Width: 640

Height: 240

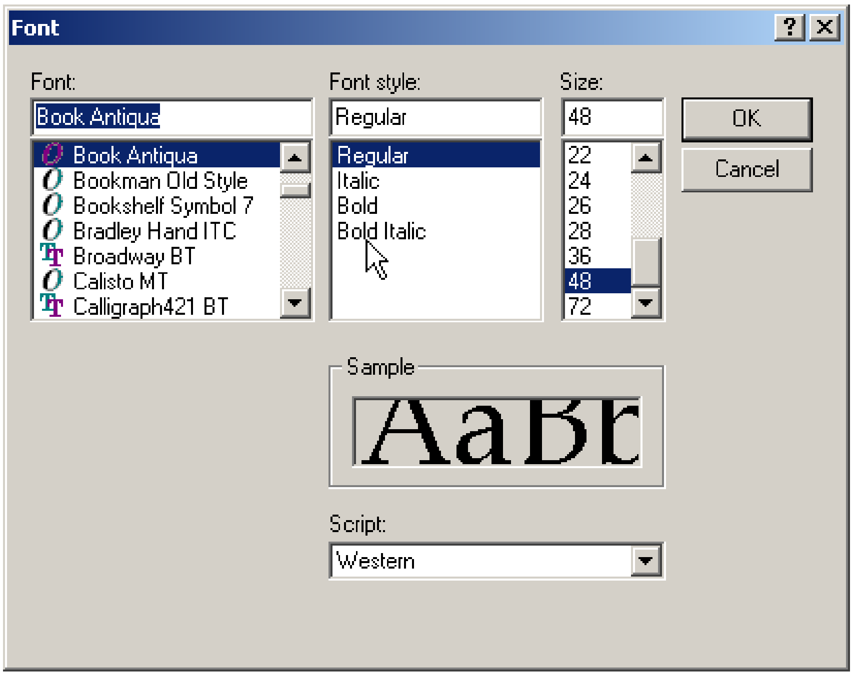
Rotation: 0

Now is the time for all good men to come to the aid of their country

Display Text Step options include:
:

- **Text to Display –** the actual text to display when the step is executed. This is Sequenceable so different text may be displayed each time this step is executed.

- **Font and Size –** these two options are configured by clicking the ‘Change’ button. This brings up the Font Settings dialog:



Font: Book Antiqua

Font style: Regular

Size: 48

Script: Western

- **Text Colour –** Clicking the ‘Change’ button next to this brings up the standard Color Selection Dialog. This is the color that the text will be displayed in. This is Sequenceable.

- **Back Colour -** Clicking the ‘Change’ button next to this brings up the standard Color Selection Dialog. This is the color for the text background. If the ‘Transparent’ checkbox is left unchecked the entire screen background will be set to this color when the step is executed.

- **Transparent –** If this is checked then the text will be displayed on top of whatever is already showing on the display. If unchecked everything on the screen is erased to the ‘Back Color’.

- **Horz Center –** This causes text to be horizontally centered in the bounding box. If unchecked then the text will be left-justified in the bounding box.

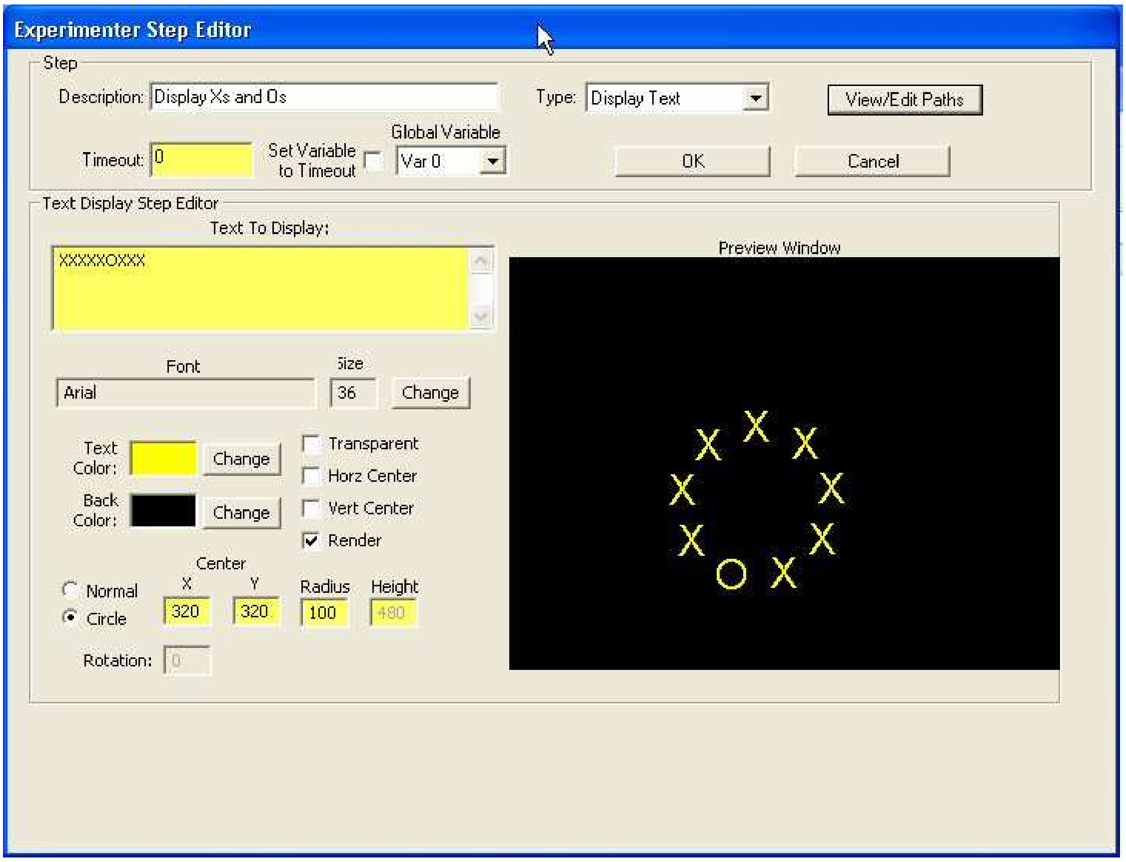
- **Vert Center -** This causes text to be vertically centered in the bounding box. If unchecked then the text will be top-justified in the bounding box.

- **Render -** If checked the text will be displayed immediately when the step executes. If unchecked the text will be drawn on the off-screen buffer and not displayed until a step such as another Display Text or a Display Image step is executed with its ‘Render’ checkbox checked. This allows layering of different images and text with different font attributes and colors for simultaneous display.

- **Normal –** When checked the text will be displayed horizontally, left to right as in the first Display Text example. Text is output to a bounding box with the upper left corner at the settings in the **Upper Left X** and **Y** edit boxes and a width and height as specified in the **Width** and **Height** edit boxes.

- **Circle –** When checked the text will be displayed in a circular pattern with the first letter at the 12 o’clock position and subsequent letters spaced evenly and placed in clockwise order. The center of the circle is at the **Center X** and **Y** edit box coordinates. The **Radius** edit box describes the radius of the circle in pixels. One good use of this setting is to place individual letters at specific screen locations by setting a Radius value of 0 and then specifying the X and Y settings to position the center of the letter precisely. Placement is always based on the center of the characters for circular output.

Example of Circular Text Output



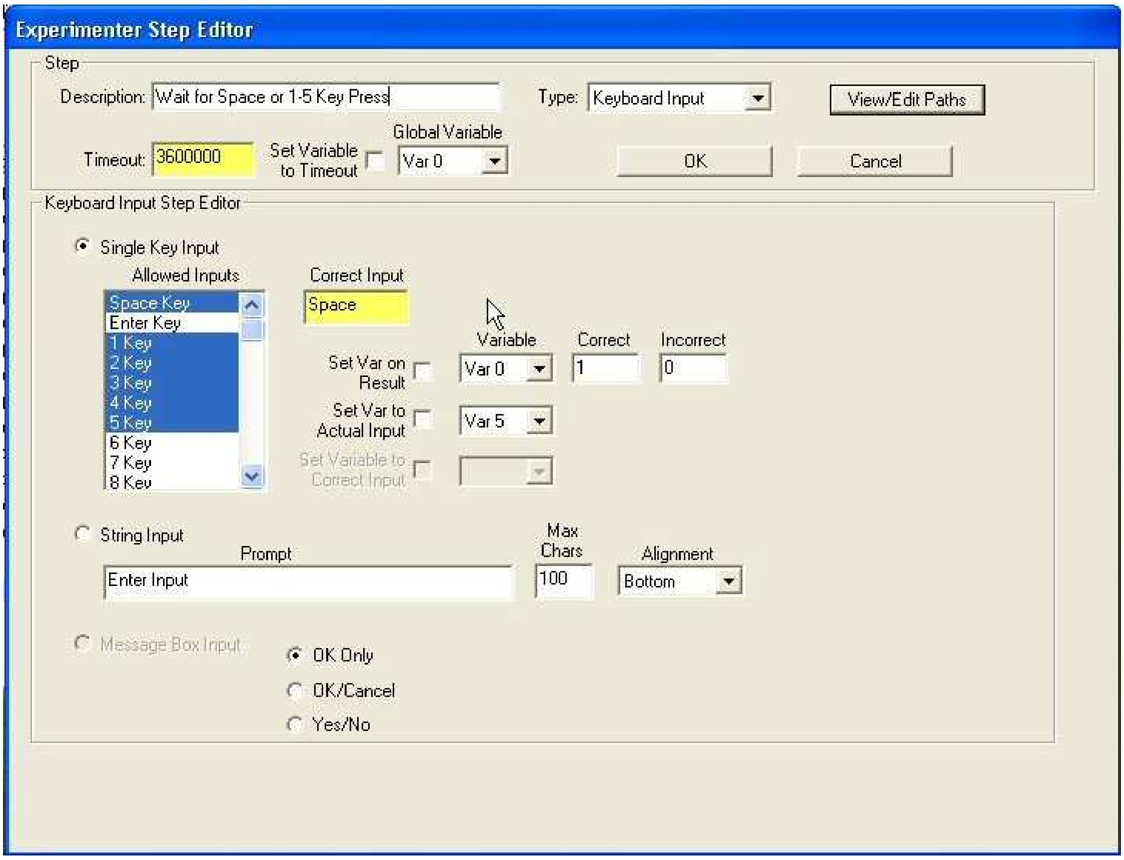
Experimenter Step Editor

Step
Description: Display Xs and Os
Type: Display Text
Timeout: 0
Set Variable to Timeout ☐
Global Variable: Var 0

Text Display Step Editor
Text To Display: XXXXXOXXX
Font: Arial
Size: 36
Text Color: 🟨
Back Color: ⬛
☐ Transparent
☐ Horz Center
☐ Vert Center
☑ Render
○ Normal
● Circle
Center X: 320
Center Y: 320
Radius: 100
Height: 480
Rotation: 0

Keyboard Input Step

The Keyboard Input Step waits for either ‘allowed’ input from the standard PC keyboard or until the value in the **Timeout** box has elapsed. Any combination of keys may be selected in the **Allowed Inputs** box using standard Windows Shift-Drag and CTRL-Click operations. When the step executes it will ignore any key pressed which is not in the **Allowed Inputs** list except for the ESCAPE key which always terminates a running experiment.



Description: Wait for Space or 1-5 Key Press
Type: Keyboard Input
Timeout: 3600000
Set Variable to Timeout ☐
Global Variable: Var 0

● Single Key Input
Allowed Inputs: Space Key, Enter Key, 1 Key, 2 Key, 3 Key, 4 Key, 5 Key, 6 Key, 7 Key, 8 Key
Correct Input: Space
Set Var on Result ☐ Variable: Var 0 Correct: 1 Incorrect: 0
Set Var to Actual Input ☐ Var 5
Set Variable to Correct Input ☐

○ String Input
Prompt: Enter Input
Max Chars: 100
Alignment: Bottom

○ Message Box Input
● OK Only
○ OK/Cancel
○ Yes/No

The above image shows a **Keyboard Input Step** in which the Space Key and Number keys 1-5 have been ‘allowed’. *Experimenter* allows most of the standard 101-key PC keyboard keys to be used and is able to differentiate between left and right Shift, Ctrl and Alt Keys.

Keyboard Input Step options include:

:

- **Single Key Input –** if checked then a single key press is expected..
- **Allowed Inputs –** a list of keys which may be used.

.

- **Correct Input –** This is a drop-down menu which allows the user to specify which key is the ‘Correct’ response. Note that this is Sequenceable so that different keys may be the ‘Correct’ key each time this step is executed.
- **Set Var On Result –** If checked then the specified Global Variable is set to the specified number in the **Correct** box if the ‘Correct Input’ matches the actual key pressed and the number in the **Incorrect** box if not a match. This Global Variable value may be used in subsequent Branch Steps and/or logged as data.

- **Set Var to Actual Input –** If checked then the specified Global Variable is set to the actual key pressed. This may be used in subsequent Branch Steps and/or logged as data.

- **String Input –** if checked then a prompt is displayed and string input is expected. This string is terminated with the Enter key or mouse click in a provided ‘OK’ button.

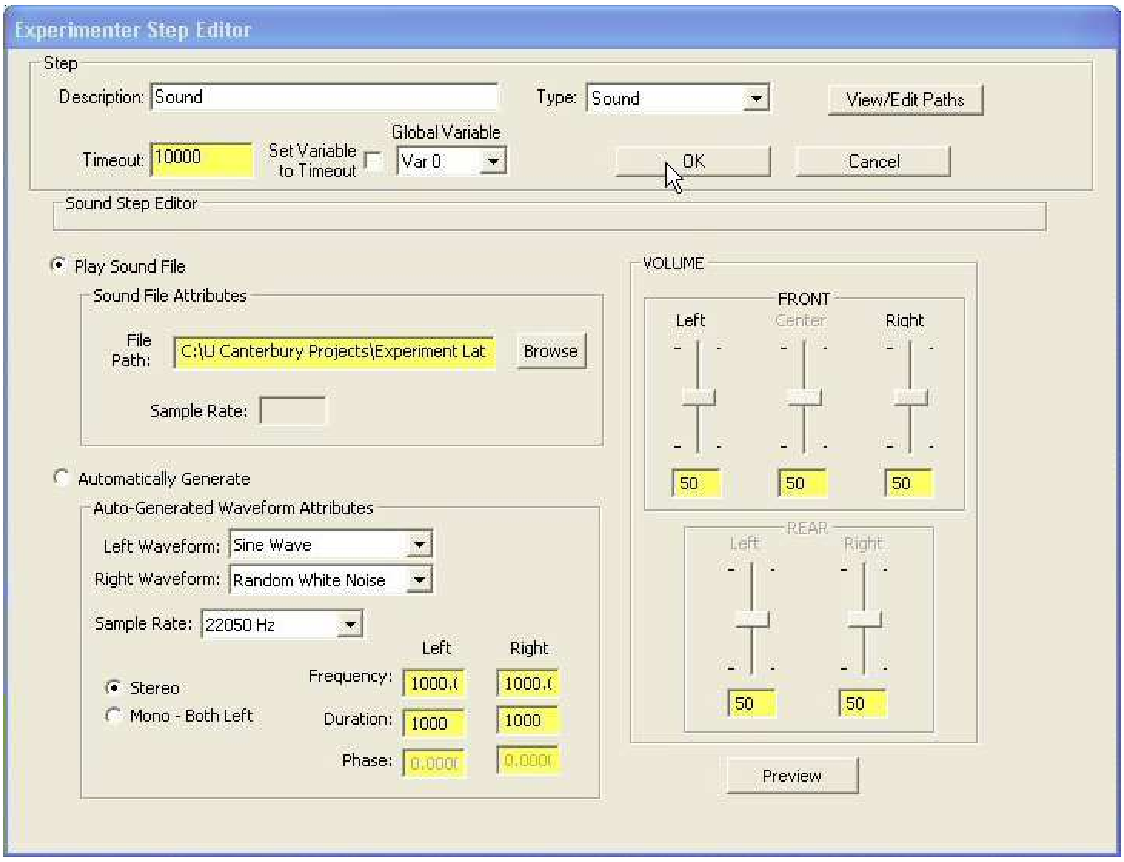
- **Prompt –** the prompt that the user sees.

- **Max Chars –** The maximum number of characters the user is allowed to type.

- **Alignment –** This is the vertical screen position for the prompt and text entry box. This may be one of Top, Bottom or Centre.

Sound Step

The **Sound Step** allows the loading and playback of pre-existing sound files in almost all of the common sound file formats. There are also a number of waveforms which can be generated on-the-fly, including: sine wave, square wave, white noise and pink noise. The frequency and amplitude for the dynamic waveforms may be specified. Any combination of sound file and dynamic waveform may be specified for either ear in stereo with any parameters.



Experimenter Step Editor

Step

Description: Sound

Type: Sound

Timeout: 10000

Set Variable to Timeout ☐

Global Variable: Var 0

Sound Step Editor

● Play Sound File

Sound File Attributes

File Path: C:\U Canterbury Projects\Experiment Lat

Sample Rate:

○ Automatically Generate

Auto-Generated Waveform Attributes

Left Waveform: Sine Wave

Right Waveform: Random White Noise

Sample Rate: 22050 Hz

● Stereo

○ Mono - Both Left

	Left	Right
Frequency:	1000.(	1000.(
Duration:	1000	1000
Phase:	0.000(	0.000(

VOLUME

FRONT	Left	Center	Right
	50	50	50

REAR	Left	Right
	50	50

Sound Step options include:

- **File Path –** Sequenceable file name for sound file to play
- **Sample Rate –** This is a read-only field showing the sample rate of the selected sound file.
- **Left/Right Waveform –** the automatically-generated waveform to use for dynamic waveforms. One value for each of two ears.
- **Sample Rate –** the sample rate to use for each channel of dynamic waveform sounds. Values range from 11,025 to 44,100 Hz.
- **Frequency Left/Right –** dynamic sound frequency. This ranges from 20 Hz to 22 kHz
- **Duration Left/Right –** length of time in milliseconds to play sounds.
- **Phase –** start of waveform. Valid values are 0.00 to 360.00 degrees.

- **Volume –** these sliders control the level of the output from 0 – 100% output.

- **Preview –** pressing this button will play the current sound(s) in the Sound Step Editor.