The Dissertation Committee for Dagmar Zeithamova Certifies that this is the approved version of the following dissertation:

CATEGORY LEARNING SYSTEMS

Committee:

W. Todd Maddox, Supervisor

Lawrence K. Cormack, Member

Alexander C. Huk, Member

Risto Miikkulainen, Member

Alison R. Preson, Member

CATEGORY LEARNING SYSTEMS

by

Dagmar Zeithamova, M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2008

Category Learning Systems

Publication No._____________

Dagmar Zeithamova, Ph.D.

The University of Texas at Austin, 2008

Supervisor: W. Todd Maddox

Category learning is an essential cognitive function. Empirical evidence and theoretical reasons suggest existence of multiple dissociable category learning systems. Here, a proposal is made that different category learning tasks are dominated by different category learning systems. A dual system theory of category learning COVIS proposes dissociation between an explicit, hypothesis-testing system, and an implicit, procedural learning system. Two studies testing this dissociation are presented, supporting the notion that hypothesis testing, utilizing working memory and explicit reasoning, mediates learning in rule-based tasks, while gradual and automatic S-R learning mediates information-integration tasks. Inconsistent findings in the literature regarding a prototype learning task suggest that two versions of this task, the A/nonA, single prototype task and the A/B, two prototype task, are mediated by distinct category learning mechanisms. A novel methodology for studying the A/nonA task and the A/B task is proposed and utilized in a functional magnetic resonance imaging study. The study reveals that the A/B task is mediated by declarative memory while the A/nonA task is mediated by perceptual learning. We conclude that at least four category learning systems exist, based on four memory systems of the brain: working memory, procedural memory, declarative memory and perceptual memory. The four category learning systems compete or cooperate during learning, each system dominating in a different category learning task. Category learning tasks provide a useful tool to understand learning and memory systems of the brain.

Table of Contents

List of Tables

List of Figures

Chapter 1: Dissociable category learning systems

INTRODUCTION

Humans live in a world of categories. Some objects or events are unique for us - *my* mother, *my* car, *my* country; but most objects we deal with are generic members of their category and we do not need to treat them uniquely. *A* cat, *a* house, *a* T-shirt - the article "a" says that we are talking about one member of a category but its individual identity is not important, not relevant for understanding the message.

Categorization is an essential cognitive function consisting in assigning entities into categories/concepts. The term “category” denotes a group of things that have something in common; the term “concept” refers to a mental representation of a category. As an essential cognitive function, categorization penetrates into every part of our lives. When we are deciding whether to take a jacket in the morning or not, we categorize the weather of the day as "jacket needed" or "jacket not needed." A doctor who looks for the best treatment for his patient needs to make a diagnosis - category assignment - and prescribe a medicine or therapy that has provided the best therapeutic effect in similar cases in the past. Social stereotypes are another form of categorization. Whenever we perceive, we perceive something as a member of a category.

Concepts and categorization serve a number of important functions, among them communication, cognitive economy, and inferences. The obvious function of categorization is providing us with a label that can be shared with others, enabling communication. However, the usefulness of the label is determined by the structure of the categories and goes well beyond communication. As proposed by Rosch (1978), concepts promote "cognitive economy." We encounter a large amount of entities every day. Without concepts, the quantity of information we would have to perceive, remember, communicate and learn about would exceed our limited cognitive capacity. Categorization divides and organizes experience into meaningful chunks. To categorize a stimulus means a) to consider it as equivalent to other stimuli in the same category, b) to consider it as different from stimuli not present in that category. The ability of a category label to provide us with information leads us to another function of categorization, its role in induction and inferences. We often use concepts when we reason about entities. Concepts are basic constituents of human thought and allow us to make inferences about properties that are not perceptible.

OVERVIEW OF THE DISSERTATION

Given the function that categorization serves in cognition, we want to know more about the process. How do we categorize new objects into known categories? How do we learn new categories? Why do we categorize the way we do? Is there a structure in natural categories? Can we reproduce it in the laboratory? What is the mental representation of categories? What neural systems support category learning? Are all categories treated equally or are there kinds of categories? In the past five years, I have been studying category learning with the hope to help answer some of these questions. In this dissertation, I will present some results of this effort. The main goal of this dissertation is twofold: first, to present evidence for dissociation between different kinds of categorization tasks; and second, to propose a connection between processes involved in these tasks and general learning and memory systems as proposed in memory literature. If valid, this connection could prove useful for both categorization research and memory research. On one hand, it would allow us to greatly increase our understanding of the cognitive process of categorization by generalizing what we know about learning and memory systems; one the other hand, it would offer memory researchers new tasks as diagnostic tools. The connections between categorization and learning and memory research has been often neglected or implicitly assumed without testing. I hope the research presented in this dissertation can help establish these connections on a more solid basis.

Chapter 1 will start with a review of theories attempting to explain human categorization and category learning. We will see that categorization has been studied under several different paradigms and explained using different theories. We will draw a conclusion that different theories may be addressing different categorization tasks and are complementary, rather than exclusive to each other. That will lead us from the question of which theory is true into a question of how different theories may be combined to fully explain human categorization behavior. We will follow with a proposal of dissociation between at least three types of categorization tasks, and a corresponding dissociation between three types of cognitive and neural processes that subserve them. The three types of tasks discussed will be: a rule-based task that is proposed to rely on an explicit hypothesis testing system; an information-integration task that is proposed to rely on a procedural learning based system; and a prototype task that is proposed to rely on a perceptual learning system. The first half of this dissertation (Chapters 1-3) will focus on the rule-based and information-integration category learning, as they afford themselves to the use of similar types of stimuli, category structures and analytic tools. The second half of this dissertation (Chapters 4-6) will focus on the prototype learning task. We will note two types of a prototype learning task that have been used in literature, propose a dissociation between them, and test it. We will conclude the dissertation in Chapter 7 with a discussion of how different kinds of categorization tasks may be supported by different cognitive and neural systems. We will revisit the outline of the dissertation later in this chapter, during the description of different categorization tasks and learning systems.

DRAWING FROM HISTORY: BASIC APPROACHES TO CATEGORIZATION AND REAL WORLD CONCEPTS

Things around us can be categorized in a number of ways. The existing human categorization systems emerged from the interaction between real-world stimuli and human cognitive processes. The study of the particular way of categorizing the world can help us to understand these processes. Cognitive psychologists have therefore aimed to reveal the inner structure of categories, to understand the process of concept acquisition, and to grasp and predict actual human categorization performance. We will first focus on categorization approaches whose emergence was primarily inspired by the aim to understand real world categories, almost at a philosophical level. Throughout the history, five major views of category learning and categorization have emerged: 1) classical, rule-based approach, 2) prototype approach, 3) exemplar approach, 4) theory-based or knowledge-based categorization, and 5) decision bound approach. We will now give an overview of the basic approaches and discuss their inspirations and possible weaknesses.

Classical view

Classical, rule-based approach began in the early twenties and was elaborated in the fifties and sixties. It proposes that humans derive a categorization rule during the process of category learning on the basis of hypothesis testing, and then use this rule when categorizing new stimuli into categories. Categories can be specified by a small number of necessary and sufficient properties. For example, *a square* is a polygon with four equal sides. When something meets these requirements, it is a member of a category, otherwise it is not. The boundary is strict.

In the early twenties, one of the first experiments addressing concept learning was Hull's experiment with “Chinese” characters (Hull, 1920). For Hull, as a behaviorist, a concept was simply the learned association between category stimuli and the category label. In his experiment, Hull used sets of cards with a “Chinese” character on it. Each set (category) had a particular pattern in common; this pattern Hull called a concept. In the learning phase, participants were to learn a label (a nonsense syllable) for a set of training cards. In the test phase, the participants were asked to novel cards with the correct concept names. Because participants were able to transfer their knowledge from training to test stimuli, they had learned to associate the labels with the defining patterns (concepts). However, when the participants were asked to draw the particular element which a card had to contain to be labeled with a particular name, it turned out that many participants were able to categorize the stimuli correctly even when they were not explicitly aware of any such common element. Hull concluded that the hypothesis testing process can occur implicitly.

In line with Hull, Vygotsky (1962) used a set of 22 cubes which differed in color, shape, height and size. On the bottom side, four non-sense words were written: "lak", "bik", "mur", "cev". During the experiment, the experimenter always showed one label at a time and asked the participants to find other cubes that might have the same word written on them. Step by step, names of other cubes were also uncovered, until the participants completed the classification of all of the cubes correctly. They should then verbalize, what characteristics "lak" had, how they differ from "bik", and so on. The height and size were the relevant dimensions for the concepts, while color and shape proved irrelevant.

Vygotsky identified three stages of category learning. In the first stage, categories are ill-formed unstructured sets. Categorization is based mainly on an impression. In the second stage, categories are complexes of concrete items, their grouping based on objective relations between the items. However, these relations are factual (as a family name), not abstract as in real concepts, or the grouping is based on a whole number of relations which have no logical connection. The third stage is one of real concept formation. Items in a category belong there because of one logically consistent characteristic (the definition of the concept) which resembles the relation objectively existing between the items. Vygotsky concluded that the basic constituent of a concept is the process of abstraction of the definition. The normative approach to what “real” concepts are is clear.

In the fifties, Bruner, Goodnow and Austin (1956) focused on the process of *concept attainment*. According to Bruner et al., concept attainment "refers to the process of finding predictive defining attributes that distinguish exemplars from nonexemplars of the class one seeks to discriminate" (p. 22). *Concept formation* then refered to sorting items into any meaningful set of classes.

The authors assumed that concepts are attained in the process of hypotheses formulations and hypotheses testing. In their experiments, Bruner et al. used a set of cards with four to six dimensions varying along two to three values. The participants were instructed as to the form of the sought-after definition (conjunctive, disjunctive, or relational). They were either shown a sequence of the cards or were allowed to choose whichever card from the set they wished. With each card they were notified whether it does or does not exemplify the intended concept. After each notification, the subjects were encouraged to make their best guess as to what the concept is. Bruner at al. were interested in the strategies the participants had used. In general, certain sub-optimalities of the hypothesis testing were observed. The adopted strategies were much more appropriate and efficient for the conjunctive than the disjunctive concepts, and the participants sought positive instances and confirmation rather than falsification.

In all the experiments that are now under the classical view, the concepts were defined by a set of necessary properties, which, when present, were sufficient for the concept membership. The concepts were expected to be acquired by forming hypotheses and their testing. In Hull's experiments, the process of hypotheses testing was rather passive and automatic – the participants were not aware they were learning a concept. In Bruner et al.'s experiments, the process was active and conscious, but as such it was forced by the experimenters' instructions. Vygotsky did acknowledge that participants went through pre-definition stages of concept formation, but concluded that the final stage of learning is a formation of a concept definition.

The main objection against the classical view has been that it does not account for most natural categories (Rosch & Mervis, 1975). Categories rarely have strict boundaries and cannot be simply defined by sets of necessary and sufficient properties. The divorce with the philosophical attempt to find such properties for at least some concepts is illustrated by a comment of Edward E. Smith (1995) that many categories (mainly the natural-kind categories) "may have necessary and sufficient conditions, but because no one knows them, they are not part of anyone's concept" (p. 29). From the perspective of Vygotsky, we would have to say that the majority of common natural concepts never reached the third stage of attainment. The attention of researchers therefore shifted to the studies of ill-defined categories and alternative models - prototype and exemplar theories - were proposed.

Prototype theory

The prototype, sometimes also called feature-based theory, was first introduced in the seventies, after several studies had shown that the classical view failed to explain some empirical phenomena observed for natural-kind and artifact categories. First, linguists, philosophers, and psychologists, who aimed to demonstrate the classical view, were unable to find definitions for the most natural concepts (E. E. Smith, 1995). Second, people disagreed with each other or even with themselves during the time, as to what is and what is not a member of a category (McCloskey & Glucksberg, 1978). Third, the classical view fails to explain typicality and prototype effect. The typicality effect means that people do not consider all members of a category as equally good members (Rosch & Mervis, 1975). For example, some birds are more birdlike than others. The prototype effect is manifested by the fact that people classify a stimulus very quickly – even when they had never perceived it before as a member of a category – if it possesses typical features shared by category exemplars or constitutes a central tendency of the presented exemplars (Posner & Keele, 1970).

The leading proponent of the prototype view of natural categories was Rosch. In a series of studies, Rosch and her colleagues (Mervis, Catlin, & Rosch, 1976; Rosch, 1975a, 1975b, 1978; Rosch & Mervis, 1975) studied several phenomena that converge and provide a measure of prototypicality: 1) the goodness of membership, assigned reliably by participants to members of a category, 2) reaction times in categorization tasks, 3) accuracy of categorization, 4) easiness of learning artificial categories, 5) free recall of the members of natural categories, and 6) possibility to substitute for a superordinate term. For some categories, such as colors or numbers, prototypes may precede the category (Rosch, 1975a). For most domains, they are abstracted from the category exemplars.

The prototype, or feature-based theory, emphasizes the family-resemblance principle proposed by Wittgenstein (1953). Wittgenstein analyzed such concepts as "game" or "tool" and argued that although there are no features common to all members of a category, the members are somewhat alike, or similar. Members of a category, like members of a family, share a number of common features, none of them necessary or sufficient. Members are generally more similar to each other than to non-members, however, the boundaries are fuzzy. In a categorization decision, a new example is compared to a prototype and if it is similar enough, it is classified as a member of the category.

But what exactly is a prototype? There are two approaches to this question. It may be either an ideal exemplar of a category, or a central tendency of a category. These two approaches are sometimes confused. Rosch considered prototypes as the clearest cases or best examples of the category, even when not central (Rosch, 1973, 1975a). For instance, categories such as "tall people, short people" may be represented by the ideals (extreme values) rather than central tendencies. The central tendency approach has been usually applied in studies of artificial categories (Posner & Keele, 1968, 1970; Reed, 1972). Recently, Goldstone (Goldstone, 1996; Goldstone, Steyvers, & Rogosky, 2003) justified the validity of both approaches and showed that categories tend to be represented by their central tendency when acquired as isolated concepts, but they tend to emphasize extreme values when acquired as contrasting concepts. Goldstone reserved the term "prototype" for the central tendency and used the term "caricature" to describe the representation that over-emphasizes the differentiating features.

Two types of stimuli are typically used in prototype learning research (Figure 1.1). One type draws from the tradition of the seminal work of Rosch (1976) and Reed (1972). In this line of research, inspired by Wittgenstein's notion of family-resemblance, the stimuli consist of sets of features, giving rise to the alternative prototype theory name, the feature-based theory. Prototypes are represented by a certain combinations of features (Figure 1.1.a) and exemplars are derived from the prototypes by altering some of the prototypical features (Figure 1.1.b). The original ambition of this type of research was to mimic internal structure of natural categories. Since then, many family-resemblance categories were constructed with the primary goal of demonstrating better fit of one categorization model against another (Murphy et al., 2005), with less concern about ecological validity of the category structures.

The second type of stimuli is dot-pattern stimuli. Prototypes are represented by a certain dot configuration (Figure 1.1.c) and category exemplars are derived from the prototypes by small changes in the position of the dots ("distortions", Figure 1.1.d). These stimuli were used in the seminal work of Posner and Keele (1968, 1970). The primary goal of their research was to use novel categories in order to address the question of how abstract representations can arise from category exemplars. The dot patterns have been used ever since as novel stimuli, without the intention to represent the structure of the real world categories.

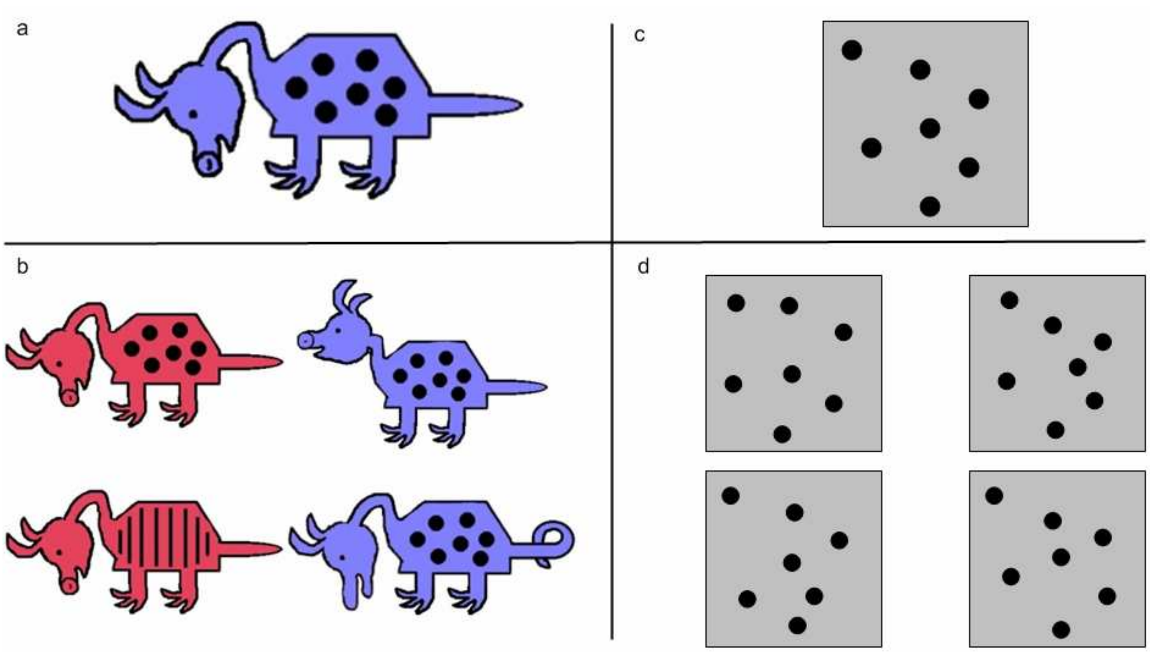

FIGURE 1.1. PROTOTYPE CATEGORY STRUCTURES. Left (a,b): Binary-valued dimension stimuli forming family-resemblance structure. a. Category prototype. b. Category exemplars. Right (c,d): Dot-pattern stimuli. c. Category prototype. d. Category exemplars.

The prototype theory has a number of limitations that prevent us from accepting it as the sole theory of either natural concepts or categorization. The central tendency (prototype) cannot be the only information abstracted in the process of category learning. There are other information people are sensitive to - size of the category, variability of its members, and within-category correlation of features (Anderson & Fincham, 1996; Chin-

Parker & Ross, 2002; Stewart & Chater, 2002). Second, not all categories have a family resemblance structure. For instance, consider the concept of a prey. To protect from predators, it is a good adaptation for a prey to be armored or to live in trees. But an animal that is both armored and lives in trees would probably not be better protected than an animal having just one of these properties. No prototype of a well-protected prey can be offered in this case. Finally, the simple prototype theory that assumes a single prototype for each category implies that a category is learnable only when it is linearly separable, but many experiments have shown that humans can learn complex, nonlinear categorization rules (Ashby & Maddox, 1992; Medin & Schwanenflugel, 1981; Shepard, Hovland, & Jenkins, 1961). The assumption of a single prototype can be however relaxed, leading to clustering and multiple prototype models (Love, Medin, & Gureckis, 2004; Verbeemen, Vanpaemel, Pattyn, Storms, & Verguts, 2007) that can account for learnability of nonlinear bounds. We will return to the prototype theory-inspired research in the second half of this dissertation in detail (Chapters 4-6).

Exemplar theory

The exemplar theory was first presented by Medin and Shaffer (1978) as an alternative to the prototype theory and hitherto remained an influential approach to categorization. Exemplar theory assumes that people represent categories by storing individual exemplars in memory. When categorizing a new object, people retrieve these exemplars and compare similarity of the new object to the exemplars of the particular category with its similarity to the exemplars of alternative categories. If the similarity to a particular category is higher than that to examples of other considerable categories, the object is classified into that category. Presentation of examples is a common way of concept learning in childhood – parents teach their children various concepts by showing instances ("look, this is a dog") or by listing instances ("examples of furniture include a chair, a table, a bed, and such things"). Prototype approach assumes that these examples only serve in the generation of an abstract concept representation - the prototype.

Exemplar approach assumes that these examples themselves are the concept representation.

One of the most widely cited exemplar models of categorization is Nosofsky's generalized context model (Nosofsky, 1986). Nosofsky accepts the geometrical model of similarity, where stimuli are represented as points in multidimensional psychological space and their similarity is a function of the distance in the psychological space. Attention may stretch or shrink the psychological space along its axes. This means that the similarity of two objects differing substantially in color and a little in shape may be high when the shape is the more attended dimension and color the less attended one, but low when attention is on the color dimension. New prototype models often adopt this geometrical model of similarity (Minda & Smith, 2001) and the geometrical model of perceptual space is also used in decision bound models (Ashby & Gott, 1988).

According to the exemplar theory, categorization of a stimulus into a category is a probabilistic function of the similarity of the stimulus to all exemplars of the given category compared to its similarity to all exemplars of alternative categories. Such a model gets over many weaknesses of the prototype theory – it accounts well for a wide variety of observed data, is sensitive to correlation information and can be successfully applied to modeling nonlinearly separable categories. Predictions based on exemplar models seem to account better for the observed results than the prototype predictors do (Nosofsky, 1987, 1992b), but this advantage may be less general than once thought. First, exemplar models are often superior to prototype models when a testing phase follows soon after learning phase. After a sufficient delay, prototype models outperform exemplar models (Reed, 1972). Several studies suggested that some kind of abstraction occurs during learning and seems itself more stable over time than are individual exemplars (Homa, 1973; Homa & Little, 1985; Posner & Keele, 1970). Also, for well-established concepts, such as “dogs”, it seems more likely that people retrieve an already pre-stored abstraction (prototype) than all encountered dog exemplars. Second, only a very limited set of category structures has been actually used to demonstrate superiority of exemplar models, mostly using categories that were ill-structured, represented by a handful of exemplars varying on small number of binary dimension (Medin & Schaffer, 1978; Medin & Smith, 1981; Nosofsky, 1987, 1992b). For categories with larger number of stimuli, prototype models outperform exemplar models (Homa, 1973; Minda & Smith, 2002; J. D. Smith & Minda, 2000). Another criticism directly addresses the central concept of both prototype and exemplar theory - similarity. Categorization of an item is supposed to be a function of its similarity either to a central tendency of a category or to stored exemplars. Rips and Collins (1993), however, argue that resemblance (similarity) is not sufficient to account for categorization. The theory-, or knowledge-based view of categories was born.

Knowledge-based view

Although the majority of this dissertation focuses on category learning when no background knowledge is available, it is important to note that alternatives to this focus have been proposed. Recall that the prototype and exemplar approaches assume that the basic factor for categorization is similarity. Things belong to the same category because they are alike. However, this notion has been questioned. An alternative, called knowledge-based (or theory-based) view, proposes that concepts are organized around personal theories about the world. These theories provide an explanation of the set of properties displayed by an instance. We group together many classes of objects on the basis of their deeper aspects. In the process of categorization, background knowledge plays a major role regardless of the surface similarity.

The role of background knowledge and participants' theories about the world was primarily demonstrated by a line of research coming from psycholinguistics (Barsalou, 1987; Labov, 1973; Lakoff, 1986; Rips, 1989; Rips & Collins, 1993) and focus primarily on categorization into concepts well known to the participants. Although something round and flat with 2 inch diameter may be more similar to a quarter than to a pizza, people still categorize it more likely to be a pizza than a quarter, because they know that quarters *cannot* be 2 inch in diameter (Rips, 1989).

Understanding the role of background knowledge and personal theories of the world in categorization is one part of the picture. To understand the other part, a divorce of the cognitive psychological tradition and the linguistics tradition is necessary. The boom of categorization research in the last three decades is characterized by a diversion from the original aim of studying real world categories and a shift to a focus on understanding category learning of novel artificial categories in carefully controlled laboratory conditions. The clear oversimplification is counterweighted by a complete control over participant's experience with category exemplars and elimination (or control) of the background knowledge. In this dissertation, we will focus on understanding category learning under these controlled conditions, studying categorization when no background knowledge is available. Although we are aware that such research offers just a part of the picture, we believe this approach enables us to gain a wealth of information about learning mechanisms that take part in complex real world concept learning.

Decision bound approach

Decision bound theory is different from the theories described above, as its current contribution to the categorization research is mainly that of methodology and a data analysis approach, rather than as a theory of category representation and categorization processes. Decision bound theory (Ashby, 1992; Ashby & Gott, 1988; Ashby & Townsend, 1986), also called general recognition theory, developed as an extension of signal detection theory (Green & Swets, 1966). The theory assumes that stimuli can be represented as points in the perceptual space.[1] The perceptual representation of a single stimulus can vary from trial to trial due to perceptual noise, just as assumed in signal detection theory. In the process of categorization, a classifier divides the perceptual space into regions and associates a category label with each region. A particular stimulus is then classified based on which region of the perceptual space it falls

[1] This assumption seems valid for perceptual but not conceptual categories (Tversky & Hutchinson, 1986), and the application of the decision bound theory has been indeed largely limited to perceptual categories.

into. The border that separates the regions is called the decision bound, analogous to the decision criterion as used in signal detection theory. In its original formulation, the decision bound theory assumed that people may directly represent decision bounds and only retrieve response label of the region (the side of the bound) that a stimulus falls into (Ashby & Maddox, 1998; Maddox & Ashby, 1993), rather than computing the similarity of a stimulus to category prototypes (as assumed by the prototype theory) or computing similarity to all category exemplars (as assumed by the exemplar theory). Currently, the concept of decision bound is used primarily as an abstract description of the character of a category structure or as a description of a participant's behavior, without implicating its ontological status. At least for some types of category structures, ontological status is however attributed to the representation of categories as regions in perceptual space and the process of categorization is assumed to involve retrieval of a response associated with that particular region of the perceptual space (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Waldron, 1999).

Coming from the signal detection framework, another important proposition of the decision bound theory was to model categories as multivariate normal distributions (Ashby & Gott, 1988, see Figure 1.2), with a potentially unlimited number of unique stimuli representing each category. This was an important step from using categories consisting of small number of exemplars varying along a few binary dimensions, as was a tradition in exemplar and prototype theory inspired research. Using the new category structures, Maddox and Ashby (1993) showed that decision bound models can outperform exemplar models. This questioned the rein of exemplar models, with their seemingly unbeatable ability to predict 90% or more of response variance (Nosofsky, 1986), and proved that there is still much unanswered in category learning research. Although originally proposed primarily for the convenience of characterizing different kinds of decision bounds and for its ability to dissociate predictions of competing categorization theories, the notion of categories as multivariate normal distributions seems to have a merit as a reasonable representation of many natural categories (Ashby, 1992; Ashby & Maddox, 1998; Flannagan, Fried, & Holyoak, 1986). The use of normally distributed stimuli as the experimental category structures and the mathematical modeling tools of the general recognition theory played a major role in further development of the category learning field. We will discuss the notions and methods of the decision bound theory in more detail as we will draw heavily from this tradition of research in the first half of this dissertation (Chapters 2-3).

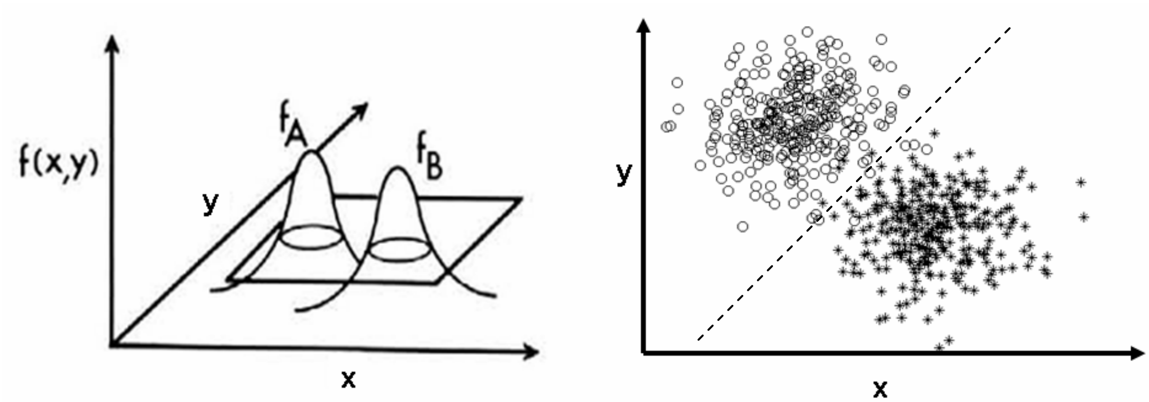

FIGURE 1.2. A SCHEMA OF CATEGORY STRUCTURES PROPOSED BY THE DECISION BOUND THEORY. Left: Categories represented by bivariate normal distributions in a two-dimensional space. x and y axis represent stimulus dimensions, f(x,y) represents probability density function for (x,y). f_A and f_B represent the probability density functions of categories A and B. From Ashby and Gott, 1988. Right: Possible categories and stimuli used in a decision bound theory experiment. Each symbol represents a stimulus with a particular value on each of two dimensions of variations (x,y). Open circles denote category A exemplars, stars denote category B exemplars. Dashed line represents the optimal decision bound.

The need for consolidation: the multiple system approach

Classical theory, prototype and exemplar theories, and the decision bound theory have all offered models of category learning that have been successful in predicting a wide array of category learning data. Researchers supporting one theory have been coming up with clever manipulations showing that their model accounts for empirical observations better than a competing model. Because all of them were successful in beating their competitors in at least some cases, no clear winner has arisen from the competition. Where does that leave us?

The first important step was to realize that rather than being based on a specific representation or process, categorization is likely based on a number of complementary processes. Both empirical evidence and our intuition tell us that in the process of categorization, people: can use rules, can abstract central tendencies, are affected by individual category exemplars, and learn associations between classes of stimuli and an appropriate response. The particular category structure used seems to play a major role in determining which categorization processes become dominant. Counterexamples are no longer used to disprove a particular theory, but rather to find conditions and limitations of it. A large body of research followed this line of thought and demonstrated that humans have available multiple mechanisms that can be used during categorization (Allen & Brooks, 1991; Erickson & Kruschke, 1998; Kemler-Nelson, 1984; Nosofsky, Palmeri, & McKinley, 1994).

The second important step was to start considering the implementation level of categorization, the neural systems that support it. Neuropsychological and neuroimaging findings provided further dissociations between different category learning tasks and suggested that many learning and memory systems of the brain are involved in category learning (Ashby & Maddox, 2005; Folstein & Van Petten, 2004; Keri, 2003; Nomura et al., 2007; E. E. Smith, Patalano, & Jonides, 1998). Originally inconsistent findings now seem to be reconcilable when we assume that different cognitive and neural mechanisms may be better suited for different kinds of categorization tasks. The first taxonomies of different categorization tasks and the corresponding learning systems have been proposed. In the next few pages, we review the proposed associations between different categorization tasks and the memory systems that subserve them.

MEMORY SYSTEMS IN CATEGORY LEARNING TASKS

Categorization tasks used in the literature can be classified in a number of ways. The stimuli used differ in the number of dimensions along which they vary and in the character of those dimensions (e.g. binary-valued or continuous). The category structures used differ in the number of exemplars they consist of, and whether they are:

overlapping or nonoverlapping, normally or nonnormally distributed, linearly separable or nonlinearly separable, probabilistic or deterministic. A well developed taxonomy of categorization tasks is currently lacking. In an extreme case, we may need a separate categorization theory for every category learning task. That could perhaps provide us with the most accurate theories, but with little understanding of the underlying principles and hard-to-generalize, compartmentalized knowledge. An alternative is to try to find a taxonomy of the existing tasks based on the processes and memory systems that they may involve. One such taxonomy has been proposed by Ashby and colleagues (Ashby & Ell, 2001; Ashby & Maddox, 2005; Ashby & Spiering, 2004). They differentiate between three types of tasks, supported by three memory systems: a rule-based task, based on working memory and reasoning; an information-integration task, based on the procedural learning system; and a prototype-distortion task, based on the perceptual learning system. We will now review the proposed task dissociations as we will use this taxonomy as the starting working hypothesis in our own investigations.

Rule-based tasks

Rule-based tasks are those in which the rule for category membership is easy to describe verbally. As such, the rule is likely to be discovered and applied by the participants in the process of learning. Examples of rule-based category structures are presented in Figure 1.3. On the left panel, an example of a rule-based category structure that uses binary-dimension stimuli is shown. The categorization rule that determines membership in category A and category B is "if the background color is blue, it is A; if the background color is yellow, it is B." The right panel shows an example of a rule-based category structure that uses normally distributed categories with continuous-value stimuli. Each point denotes a stimulus, for instance a Gabor patch varying in spatial frequency and orientation.[2] The categorization rule that determines membership in category A and category B is "if the spatial frequency is low, it is A; if the spatial

[2] Gabor patch is a sinewave grading enclosed in a Gaussian envelope (see examples on Figure 1.3, right panel)

frequency is high, it is B." Note that the word "rule" is used here in two different meanings. First, for most category structure, there is some formal description of how category membership can be determined. This description is often called a categorization rule, no matter of its character. However, the category structures are only called rule-based if this categorization rule is likely to be explicitly learned by the participants. Most unidimensional rules involving separable dimensions do constitute rule-based categories; most rules that are of higher level complexity than a conjunctive rule do not (Ashby & Maddox, 2005).

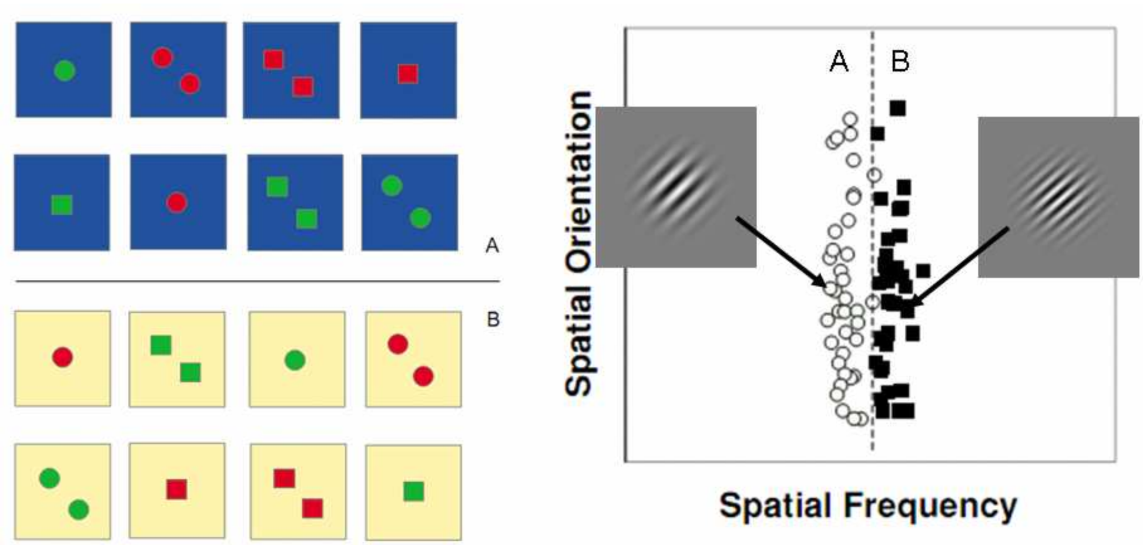

FIGURE 1.3. RULE-BASED CATEGORY STRUCTURES. Left: Stimuli varying along four binary-dimensions. The categorization rule is "blue background stimuli belong to category A, yellow background stimuli belong to category B." From Ashby and Ell, 2001. © 2008 Elsevier Ltd. All rights reserved. Reprinted with permission. Right: Normally distributed categories with large number of exemplars and stimuli varying along two continuous dimensions. Each point denotes a stimulus (Gabor patch), two example stimuli are presented. The categorization rule is "low spatial frequency stimuli are category A, high spatial frequency stimuli are category B."

Although on the surface, the two tasks presented in Figure 1.3 may look quite different for a participant, they both can be learned by the same hypothesis testing process. The classical approach to categorization (Bruner, Goodnow, & Austin, 1956) and most multiple systems approaches postulate a categorization system that is based on such extraction of a categorization rule via hypothesis testing and explicit reasoning (e.g. Anderson & Betz, 2001; Ashby & Ell, 2001; Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994; Patalano, Smith, Jonides, & Koeppe, 2001; E. E. Smith, Patalano, & Jonides, 1998). The hypothesis testing system involves working memory and attentional processes and is thought to rely on the prefrontal working memory system (Ashby & O'Brien, 2005; Tracy et al., 2003).[3]

Working memory and the prefrontal system in rule-based learning

Working memory is a short-term ability to maintain and manipulate limited amount of information (Baddeley, 1995). Prefrontal cortex has been implicated as the crucial neural structure supporting working memory and attention (D'Esposito & Postle, 1999; D'Esposito, Postle, Stuss, & Knight, 2002; Goldman-Rakic, 1987, 1990; Narayanan et al., 2005), with anterior striatum, reciprocally connected to the prefrontal cortex, being part of the working memory network (Hikosaka, Sakamoto, & Usui, 1989; R. Levy, Friedman, Davachi, & Goldman-Rakic, 1997; Schultz et al., 1995).

Although working memory is a limited time memory, not a long-term memory, it can mediate learning in a rule-based task since the task has a structure simple enough to be discovered quickly by a reasoning process (Ashby & O'Brien, 2005; E. E. Smith & Grossman, 2008). Limited working memory capacity determines the complexity of stimuli and rules that can be learned. The role of the prefrontal cortex in rule-based learning has been reported by several neuroimaging studies (E. E. Smith, Patalano, & Jonides, 1998; Tracy et al., 2003), and also seems firmly established in the neuropsychological research as the Wisconsin card sorting test (a rule-based task) is widely used for neuropsychological diagnosis of frontal lesions (Robinson, Heaton, Lehman, & Stilson, 1980). The notion of the working memory basis of rule-based learning is more recent (Ashby & O'Brien, 2005; E. E. Smith & Grossman, 2008) and has

[3] Other names used for the hypothesis testing system are: *verbal* system, reflecting its role in learning of verbalizable rules; *explicit* system, reflecting the explicit, conscious awareness accessible nature; and *rule-based* system, somewhat tautological name reflecting that it is a mechanism for learning *rules*.

been supported by only a few studies. The goal of the first half of the dissertation is to increase our understanding of the working memory role in category learning. We will return to a theory of rule-based and information-integration learning later in this chapter. Chapter 2 and Chapter 3 will then provide behavioral evidence for the preferential role of working memory in rule-based learning.

Information-integration tasks

Information-integration tasks are those tasks in which information from multiple dimensions need to be integrated at some pre-decisional stage. This can be done, for instance, by a linear combination of dimensional values or by treating each stimulus as a gestalt rather than analyzing it into its constituent components (Ashby, Ell, & Waldron, 2003). Figure 1.4 shows two examples of information-integration category structures, using the same stimuli as the rule-based structures to clarify the differences. On the left panel, an information-integration category structure with stimuli varying along four binary dimensions is presented. The rule for category membership is not immediately obvious: “if a stimulus has at least two out of the three following features - blue background, two embedded symbols, embedded symbols of green color - it is A; otherwise, it is B.” Recall the dual meaning of the word “rule”, with its narrow meaning in the context of “rule-based structure.” Although there *exists* a rule that separates category A exemplars from category B exemplars, this rule is rather complex, and as such is not likely to be explicitly discovered, represented and applied by participants in this form. The category structure is thus not considered rule-based. The right panel presents a category structure based on bivariate normally distributed stimuli. The rule that determines category membership is “if orientation is bigger than spatial frequency, it is A; if spatial frequency is bigger than orientation, it is B.” Again, although an optimal rule exists, it is not likely to be discovered and explicitly used by the participant as it requires combination of incommensurable units.

The use of the category structure on right panel is much more popular and will be primarily used here because using alternative, non-information-integration strategies is made difficult. The large number of unique stimuli makes memorization inefficient and the elongation of the categories along the decision bound ensures that unidimensional rules provide poor accuracy (compare with the right panel on Figure 1.2 where the optimal bound location is the same, but a unidimensional rule (some decision bound perpendicular to one of the axes) could yield accuracy only marginally worse). The category structure on the left panel can be solved using a number of strategies, among them exemplar memorization or use of a rule-plus-exceptions verbal strategy (Erickson & Kruschke, 1998) being most common.

Another popular information-integration task is the weather prediction task in which participants need to integrate information from several probabilistic cues in order to predict which of two outcomes (rain or shine) is likely to occur. Although we will include results of the weather prediction task in the discussion of the information-integration task, it is important to note that participants in this task are often found to use a mix of non-information integration as well. Common is the use of unidimensional rules, especially early in the learning (Gluck, Shohamy, & Myers, 2002) and memorization of the limited set of stimuli (14), especially later in the learning (Knowlton, Squire, & Gluck, 1994).[4]

[4] Note also that most prototype-based category structures require integration of information across several stimulus dimensions and thus can be (and sometimes are) viewed as a subtype of the information-integration structures. However, it seems that the prototype structures promote additional learning mechanisms, and we will thus consider them separately.

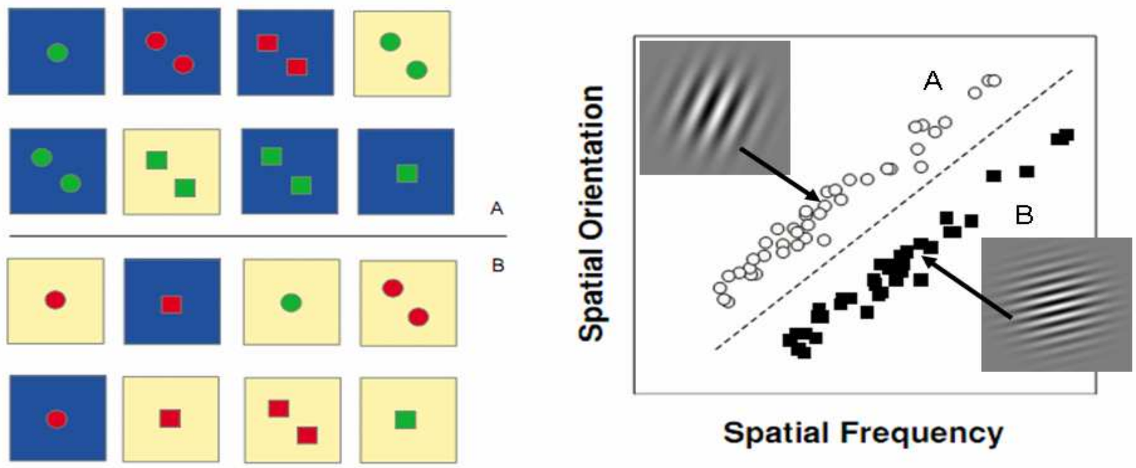

FIGURE 1.4. INFORMATION-INTEGRATION CATEGORY STRUCTURES. Left: Stimuli varying along four binary-value dimensions. From Ashby & Ell, 2001. Right: Normally distributed categories with stimuli varying along two continuous dimensions. No simple verbalizable rule separates category A stimuli and category B stimuli.

In the current view, it is assumed that participants in the information-integration task learn the associations between regions in the perceptual space (groups of stimuli) and the category label, rather than deriving any explicit categorization rules. The mechanism of learning is thought be implicit, procedural learning-based and dependent on the posterior striatum (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Waldron, 1999; Nomura et al., 2007).

Procedural memory and the striatum in information-integration learning

Procedural memory, also called the *habit* memory, is a form of nondeclarative memory. It is memory for "how to" – how to ride a bike, ski, play tennis, play the piano, etc. The characteristic feature of procedural memory is that it is acquired gradually through practice and is not easily communicated verbally to others (e.g., try to describe how you tie shoe laces). It is now studied in the context of acquisition of both motor skills and cognitive skills. Popular paradigms for studying procedural learning are, for example, a serial reaction time task (Nissen & Bullemer, 1987; Reber & Squire, 1994, 1999; Willingham, Nissen, & Bullemer, 1989) and control of complex systems (Berry &

Broadbent, 1984, 1988). The key neural structure in many kinds of procedural learning is the striatum (Curran, 1995; Mishkin, Malamut, & Bachevalier, 1984; Westwater, McDowall, Siegert, Mossman, & Abernethy, 1998).

In the context of category learning, the procedural system is assumed to gradually and incrementally learn to associate specific regions in the perceptual space with category assignments (Ashby & Waldron, 1999; Knowlton, Mangels, & Squire, 1996). The proposed circuit (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Waldron, 1999) starts with stimulus representation in inferotemporal cortex. Many-to-one convergence of inferotemporal cells onto the posterior caudate (Wilson, 1995) creates a low-resolution representation of the perceptual space within the caudate. The striatum then functions to associate a pattern of cortical activation with a motor response, by strengthening recently activated synapses after dopamine-mediated reward (Wickens, 1993). The striatum has been implicated in linking a stimulus with classification response in both human and animal research (for a review of this research, see Packard & Knowlton, 2002). Patients with striatal damage are impaired on information-integration tasks (Filoteo, Maddox, & Davis, 2001a; Knowlton, Mangels, & Squire, 1996; Knowlton, Squire, Paulsen, Swerdlow, & Swenson, 1996), but amnesiac patients are not (Eldridge, Masterman, & Knowlton, 2002; Filoteo, Maddox, & Davis, 2001b; Knowlton, Squire, & Gluck, 1994).

Dissociation between rule-based and information-integration category learning

Ashby and colleagues (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Waldron, 1999) proposed a formal model called COVIS (COmpetition between Verbal and Implicit Systems), describing how the hypothesis-testing system and the procedural system interact in the process of category learning. They proposed that both learning systems attempt to acquire and solve every categorization task encountered. However, the relative weight of each system in the category judgment depends on the relative success of each system in category learning, which in turn depends on the type of category structure to be acquired. The dissociation between rule-based and information-integration categorization tasks comes from the tradition of decision bound theory. The nature of the decision bound – whether it is easy to describe verbally or not – discriminates the rule-based structures from the information-integration structures. The hypothesis-testing system searches for and applies explicit rules that are easy to verbalize; it thus dominates in learning of rule-based category structures. The procedural learning-based system learns to associate a category response with a region of perceptual space without deriving any explicit rule. Although slower, the procedural system dominates learning in the information-integration category structures because no simple rule is likely to be discovered by the hypothesis-testing system. In contrast, category structures acquired by the procedural system may be very complex (e.g. Ashby & Maddox, 1992, 2005).

As a consequence of the proposed learning mechanisms and the underlying neurobiology of the two systems, manipulations that affect one but not the other system should differentially affect rule-based and information-integration learning. This has been confirmed in a number of studies (see Maddox and Ashby, 2004 for a review). Several experiments have introduced manipulations that affect information-integration, but not rule-based category learning. First, because feedback-mediated dopamine release is thought to play a crucial role in strengthening the cortico-striatal synapses mediating the stimulus-response associations, timely feedback should be crucial for information-integration learning. When the feedback is delayed, the recently activated synapses may return to baseline and the stimulus-response association may not be strengthened (Arbuthnott, Ingham, & Wickens, 2000; Kerr, & Wickens, 2001). Maddox and colleagues (Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005) indeed showed that with a five second delay between a response and feedback, learning of an information-integration category structure becomes impossible, while learning of a rule-based structure is minimally affected. Second, the procedural component of information-integration learning should be critically dependent on consistent stimulus-response mapping. Most category learning experiments use consistent stimulus-response mappings. In a typical paradigm, the participant is asked to press button "A" with the left hand and button “B” with the right hand to indicate category A and category B stimulus (A-B training). Maddox, Bohil and Ing (2004; see also Ashby, Ell, & Waldron, 2003) used a variable stimulus-response mapping. The participants were asked to press either button “Yes” or button “No” to a stimulus in response to a question “Is this an A?” or “Is this a B?” (Yes-No training). As the stimulus-response-outcome mapping was inconsistent, the Yes-No training impaired information-integration category learning compared to A-B training, but had no effect on the rule-based category learning.

The previous studies show that the procedural system differs from the hypothesis-testing system in that it requires timely feedback and a consistent stimulus-response mapping. When these are not provided, learning by the implicit system is adversely affected. These manipulations have minimal effect on rule-based learning because the hypothesis-testing system utilizes consciously accessible working memory that can hold the feedback information over the delay and uses explicitly represented categorization rule that can be flexibly tested and applied with both A-B training and Yes-No training.

One may argue that the disruption of information-integration category learning but not rule-based learning in the previous studies is due to differences in complexity and therefore difficulty of simple (usually one-dimensional) verbalizable rules in the rule-based condition versus complex nonverbalizable rules in the information-integration condition. To provide evidence for the existence of two alternative systems, double dissociation should be demonstrated. A manipulation that affects working memory load should affect the hypothesis-testing, but not procedural learning system. Chapter 2 and Chapter 3 are dedicated to discussing and demonstrating this dissociation.

Prototype learning tasks

Prototype learning tasks, inspired by the prototype theory, represent categories as collections of stimuli that are generated from a single prototype by various alterations of the prototypical values. Category exemplars are often called “prototype distortions”. Recall the two examples of prototype-based category structures presented in Figure 1.1. The left panels (Figure 1.1.a,b) show a prototype and a collection of category exemplars, using binary-valued stimuli. These types of categories represent the traditional family-resemblance based natural categories as proposed by Wittgenstein (Wittgenstein, 1953) and promoted by Rosch (Rosch & Mervis, 1975). The right panels (Figure 1.1.c,d) show a prototype and a collection of category exemplars using dot-pattern stimuli. One pattern of dots is the prototype, and category exemplars are created by randomly moving the location of each dot in the prototype. These types of categories follow the research of prototype theorists who focused on prototype abstraction in novel artificial categories (Posner & Keele, 1968, 1970; Reed, 1972).

The prototype can be learned by extracting the common features or the common structure from the category exemplars. It is thought that this extraction depends on the perceptual representation memory mediated by learning-related changes in the visual cortex (Ashby & Casale, 2003; Casale & Ashby, in press; Reber & Squire, 1999).

Perceptual memory and the cortical system in prototype learning

The perceptual representation memory system is a form of implicit (nondeclarative) memory, manifested by an improvement in the perception and processing of a repeated stimulus (Dosher & Lu, 1999; Schacter, 1990, 1994; Tulving & Schacter, 1990). Perceptual memory has been primarily studied in the context of perceptual repetition priming (Schacter, 1994; Schacter, Cooper, & Delaney, 1990), a phenomenon of speeded identification of a recently encountered stimulus even in the absence of explicit memory for the prior exposure. Repetition priming has been shown to be independent of declarative (medial temporal lobe-based) memory (Schacter, Cooper, & Delaney, 1990; Schacter, Cooper, Tharan, & Rubens, 1991), but also independent of nondeclarative striatal memory (Knowlton, Squire, Paulsen, Swerdlow, & Swenson, 1996; Reber & Squire, 1999). Priming and the perceptual memory are mediated by stimulus related changes within the sensory cortex, primarily by a reduction of activity for repeated stimuli (Schacter & Buckner, 1998; Schacter, Wig, & Stevens, 2007; Slotnick & Schacter, 2006; Wiggs & Martin, 1998).

Perceptual memory can be elicited not only for identical, but also perceptually similar stimuli (Biederman & Cooper, 1992; Cooper, Schacter, Ballesteros, & Moore, 1992; Wagner, Gabrieli, & Verfaellie, 1997). Perceptual memory seems thus ideally suited for prototype extraction from a set of category exemplars as all category stimuli are typically perceptually similar. Keri and colleagues (Keri et al., 2002) proposed a computational model of prototype learning based on fast synaptic changes of lateral connections within the sensory cortex, following a Hebbian rule. These synaptic changes can support prototype formation in a self-organizing manner, without external feedback. The neurobiological plausibility of this model is supported by an observation that such fast synaptic changes occur in the primary visual cortex of rat (Varela et al., 1997). Priming-like changes in the visual cortex were indeed found in several fMRI studies of prototype learning (Aizenstein et al., 2000; Reber, Stark, & Squire, 1998a, 1998b) and steep typicality gradients in prototype learning emphasize the role of perceptual similarity for the perceptual memory system (Casale & Ashby, in press; J. D. Smith & Minda, 2002). We will return to prototype learning in the second half of the dissertation in detail (Chapters 4-6).

And what about declarative memory? Completing the picture

In the quest for understanding category learning, we have arrived at the conclusion that different category learning tasks seem to be tapping into different learning and memory systems. The taxonomy dissociating rule-based, information-integration, and prototype category structures, dependent on working memory, procedural memory, and perceptual representation memory respectively (Ashby & Ell, 2001; Ashby & Maddox, 2005; Ashby & Spiering, 2004) seems well supported and is covering the majority of the category structures used. However, one important memory system – the explicit, declarative memory system mediated by the medial temporal lobe – has not been mentioned. Categorization tasks are often considered non-declarative, hippocampus-independent forms of learning. Indeed, many neuropsychological studies have contrasted category learning with recognition in amnesic patients and found spared category learning in rule-based Wisconsin Card Sorting Test (Janowsky, Shimamura, Kritchevsky, & Squire, 1989; Leng & Parkin, 1988), in an information-integration task with large categories (Filoteo, Maddox, & Davis, 2001b), in the weather prediction task (Knowlton, Mangels, & Squire, 1996; Knowlton, Squire, & Gluck, 1994) and in the prototype learning task (Bozoki, Grossman, & Smith, 2006; Knowlton & Squire, 1993). Based on these findings, Ashby and Waldron (2000) argued that medial temporal lobe structures are not critical for most forms of category learning. On the opposite side, exemplar models assuming declarative memory keep providing good fits to a range of categorization data (Nosofsky, 1992a; Nosofsky & Johansen, 2000; Shin & Nosofsky, 1992; Stanton, Nosofsky, & Zaki, 2002). Nosofsky and Zaki (1998) argued that all category learning is ultimately dependent on exemplar memory mediated by the medial temporal lobe, and the seeming dissociation between recognition and categorization is solely due to lower sensitivity of the categorization tasks compared to the recognition tasks. So, is there a role for declarative memory in category learning?

Probably neither extreme is correct. First, even when the neuropsychological findings indicate that an intact medial temporal lobe is not necessary for many categorization tasks, the current view is that exemplar memory plays at least a complementary role in many category learning tasks in normal participants (Ashby & Ell, 2001; J. D. Smith & Minda, 2001). Knowlton (1999) in response to Nosofsky and Zaki acknowledged the role of (declarative) exemplar memory besides other category learning mechanisms, but suggested that exemplar knowledge acquisition and category knowledge acquisition should not be considered as mutually exclusive mechanisms of category learning. Rather, future research should focus on the circumstances under which one or the other is more likely to occur. In line with this proposal, many experiments focused on identifying conditions suited for exemplar memorization. They identified its preferential role when categories are small, consisting of only a few exemplars (Homa, Sterling, & Trepel, 1981; Minda & Smith, 2001), when categories are poorly structured (Lei & Zhansheng, 2003; J. D. Smith & Minda, 2000), when individual exemplars are repeatedly presented (Knowlton, Squire, & Gluck, 1994; J. D. Smith & Minda, 1998) and when an exemplar constitutes a salient exception to the categorization rule (Erickson & Kruschke, 1998; Nosofsky & Palmeri, 1998; Nosofsky, Palmeri, & McKinley, 1994).

The paragraph above suggests that there is a declarative component to many categorization tasks. However, is there a categorization task that may be *primarily* dependent on the declarative memory? Recently, a suspicion turned surprisingly to one version of the prototype task. A closer look at the literature shows that two versions of the prototype tasks exist, A/nonA version and A/B version. In the A/nonA version, only one category exists, consisting of distortions from a single prototype, and the participants need to distinguish categorical (A) from noncategorical (nonA) items. In the A/B version, two categories exist, derived from two distinct prototypes, and the participants need to distinguish category A exemplars from category B exemplars. Ashby and Casale (2003) discussed the role of the perceptual representation system in prototype learning and pointed out that while a feeling of perceptual familiarity (mediated by the perceptual memory) can support performance of the A/nonA task, it is not, by itself, sufficient in the A/B task, as both A exemplars and B exemplars would elicit familiarity. Zaki and colleagues (Zaki, Nosofsky, Jessup, & Unversagt, 2003) tested both A/B and A/nonA prototype task in amnesiac patients and found intact A/nonA learning, but impaired A/B learning. Based on this finding, Ashby and colleagues (Ashby & Maddox, 2005; Ashby & O'Brien, 2005) proposed that while the A/nonA prototype task may be mediated by perceptual memory, the A/B prototype task may be mediated by explicit, declarative memory. The second half of this dissertation is dedicated to examination of this notion (Chapters 4-6).

Chapter 2: Dual task interference in category learning[5]

In Chapter 1, we discussed the COVIS model proposed by Ashby and colleagues (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Waldron, 2000). COVIS postulates two systems that compete throughout learning – a hypothesis-testing system that uses logical reasoning and depends on working memory and selective attention, and a procedural learning-based system, that gradually learns associations between regions in the perceptual space and a category label through practice, without deriving explicit rules. COVIS proposes that the hypothesis-testing system dominates learning of rule-based category structures while the procedural system dominates learning of information-integration category structures. Behavioral dissociation between the hypothesis-testing system and the procedural learning system can thus be demonstrated by manipulations that selectively affect rule-based learning or information-integration learning.

Recall that several studies demonstrated manipulations that affected information-integration, but not rule-based learning using delayed feedback or inconsistent stimulus-response mapping (Ashby, Ell, & Waldron, 2003; Maddox, Ashby, & Bohil, 2003; Maddox, Bohil, & Ing, 2004; Maddox & Ing, 2005). One may argue that information-integration is more complex than rule-based learning, so the reverse dissociation needs to be demonstrated as well. Recent studies began to explore conditions that affect rule-based, but not information-integration learning using working memory load manipulations (Maddox, Ashby, Ing, & Pickering, 2004; Maddox, Filoteo, Hejl, & Ing, 2004). Waldron and Ashby (2001) introduced a concurrent, attention demanding task that should affect the hypothesis-testing, but not procedural system. They used binary-valued stimuli and found large dual task interference on uni-dimensional rule-based learning and small dual task interference on (multidimensional) information-integration learning. The authors concluded that their study demonstrated the role of the hypothesis testing system

[5] Major portions of this chapter have been previously published as an article Zeithamova & Maddox (2006). Memory & Cognition, 34(2), 387-398. Copyright 2006 Psychonomic Society, Inc

and working memory in rule-based learning and supported the multiple system view of categorization.

In this chapter, we explore the role of a dual task in category learning in more detail. We begin by reviewing the study of Waldron and Ashby (2001), noting a few possible shortcomings and presenting an alternative explanation of their results suggested by Nosofsky and Kruschke (2002). We then present two experiments offering additional evidence for the dissociation between the rule-based and information-integration category learning. In the first experiment, we test and extend the generality of Waldron and Ashby's results when applied to a uni-dimensional rule-based and (two-dimensional) information-integration category learning task using a large number of perceptually similar continuous-value dimension stimuli. Second, we examine the dual task interference in two-dimensional, conjunctive rule-based learning and provide a critical test of Nosofsky and Kruschke's single system explanation of the original results.

REVIEW OF WALDRON AND ASHBY (2001)

Recall that COVIS postulates that the hypothesis-testing system relies on working memory and selective attention to solve rule-based category tasks, whereas learning in the procedural-learning system is essentially automatic. Waldron and Ashby (2001) provided an empirical test of this prediction by comparing rule-based and information-integration category learning under dual task conditions with that in a single-task control. They chose a numerical analog of the Stroop task (for a detailed review of the Stroop task, see MacLeod, 1991) to serve as a dual task. The Stroop task is known to require working memory and selective attention, and to strongly activate the anterior cingulate and prefrontal cortex (Bench, Frith, Grasby, & Friston, 1993), neural structures associated with the explicit, hypothesis-testing system, but not with the implicit procedural-learning system proposed in COVIS.

Waldron and Ashby had participants learn to categorize stimuli that varied on four binary dimensions (see the left panels of Figures 1.3 and 1.4 in Chapter 1). In the uni-dimensional rule-based condition, one dimension was relevant and the remaining three were irrelevant (Figure 1.3, left panel). In the information-integration condition, information from three dimensions had to be integrated and one dimension could be ignored (Figure 1.4, left panel). Under the control conditions, the participant simply categorized each stimulus on every trial and received feedback after each response. In the dual task conditions, the participant had to perform a numerical analog of the Stroop task during each trial of categorization. The Stroop task stimulus was presented simultaneously with the categorization stimulus for 200 ms. The Stroop stimulus was then masked and the categorization stimulus remained on the screen until the participant categorized it. After categorization feedback, the participant was to respond to the Stroop stimulus they had seen at the beginning of the trial. Therefore, the participant was required to hold a representation of the Stroop stimulus in the working memory during the process of categorization and while receiving feedback. Performance in the Stroop task was emphasized over the categorization task. Participants were tested on four rule-based tasks and four information-integration tasks. Participants performed each task either until they reached the learning criterion of 8 trials in a row correct or until they reached 200 trials.

Waldron and Ashby found that the dual Stroop task produced greater interference for the unidimensional rule-based task than for the information-integration task (Figure 2.1). These findings support the COVIS prediction that the working memory and attention requiring hypothesis testing system supports rule-based learning, while the (automatic) procedural system supports information-integration learning.

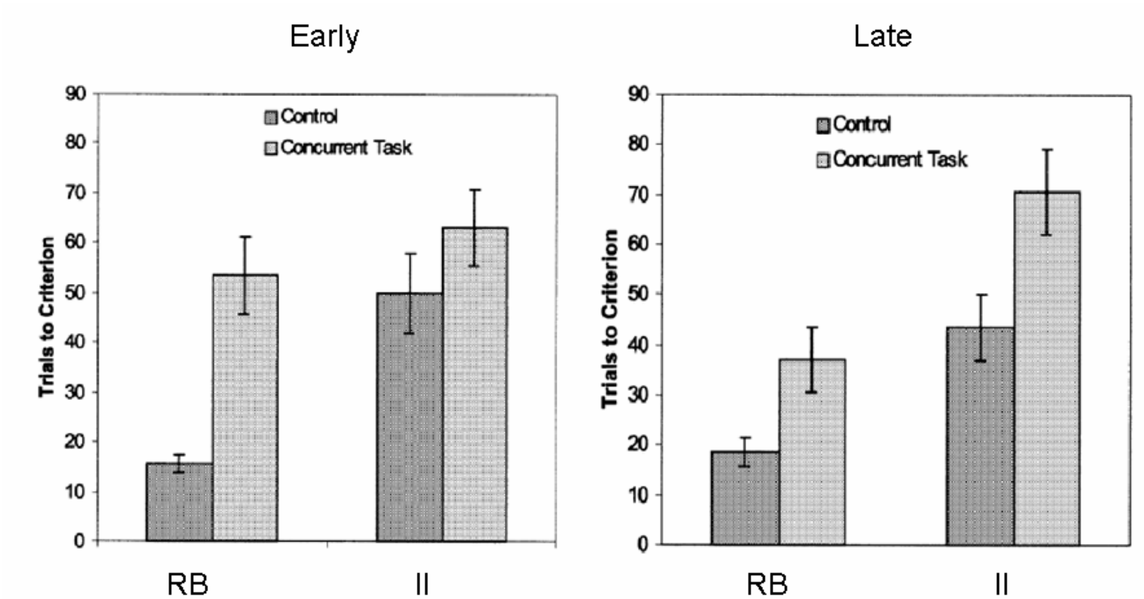

FIGURE 2.1. RESULTS FROM WALDRON AND ASHBY (2001). Early: Results from the first half of the experiment. Late: Results from the second half of the experiment. RB: rule-based; II: information-integration. Copyright 2001 Psychonomic Society, Inc. Reprinted with permission.

We found the results compelling, but a few possible weaknesses need to be noted. First, examine the results presented on Figure 2.1. The category structure x condition interaction was present only during the first half of the experiment. Additionally, because of the different baseline performance, the smaller dual task effect on information-integration learning could be a result of the floor effect. Second, the eight correct responses in a row learning criterion might be too lenient. We conducted a series of simulations with the original Waldron and Ashby stimulus sequences (in these sequences each stimulus was presented once in each block of 16 trials) and found that the eight-in-a-row criterion could be reached through random responding within two hundred trials with probability .33. Also, a participant using a unidimensional rule to solve the information-integration tasks could respond correctly on each individual trial with probability .75. Third, as we discussed in Chapter 1, exemplar memorization is likely to be operative with so few stimuli. This may have involved different neural structures than the authors assumed and thus may have influenced the results in an unknown way. The goal of

Experiment 1 is test the role of working memory in rule-based learning using the dual task methodology while avoiding these possible shortcomings.

A different criticism of Waldron and Ashby's experiment came from Nosofsky and Kruschke (2002), who questioned its interpretation. They claimed that a single system exemplar model can account for the larger dual task interference in the rule-based task when a selective attention parameter (such as proposed in Kruschke, 1992) is varied between the control and the dual (Stroop) condition. As only one dimension is relevant in the rule-based task, the selective attention requirement for good performance is high. Selective attention requirement is low in the information-integration task, as attention can be spread across all dimensions without affecting performance. Experiment 2 tests this notion.

EXPERIMENT 1

The main aim of Experiment 1 was to test the generalizability of Waldron and Ashby's (2001) results while avoiding their possible shortcomings, using a large number of unique stimuli varying along continuous-valued dimensions. The stimuli were Gabor patches that varied across trials in spatial frequency and spatial orientation. Unidimensional (UD) rule-based and information-integration (II) category learning were examined under control and dual Stroop conditions. The scatter plots of the stimuli used in the UD and II category learning conditions are shown in Figure 2.2 along with the optimal decision bound. Each point in the scatterplot denotes the spatial frequency and spatial orientation of a single stimulus. In the unidimensional condition, spatial frequency was the relevant and spatial orientation was the irrelevant dimension. The optimal rule required participants to respond A when the spatial frequency was low and to respond B when the spatial frequency was high. Both dimensions were relevant in the information-integration condition. The optimal rule required participants to respond A when the difference of the value on spatial frequency dimension and the value on the spatial orientation dimension was low and to respond B when the difference of the values on the two dimensions was high. Such a rule is not likely to be explicitly learned and verbalized by the participants because it compares values in different units. The category discriminabilities (d') in the physical space were 4.5 for unidimensional and 10.3 for information-integration category structure.[6]

Method

Participants

One hundred seventy students at The University of Texas at Austin participated in the experiment in partial fulfillment of a class requirement or for pay. All observers were tested for 20/20 vision, and no observer completed more than one experimental condition. Each participant completed one of four experimental conditions: unidimensional rule-based, control (UDC), unidimensional rule-based, dual Stroop (UDS), information-integration, control (IIC), and information-integration, dual Stroop (IIS).

Stimuli and apparatus

The categorization stimuli were Gabor patches that varied across trials in spatial frequency and spatial orientation. The experiment used the randomization technique introduced by Ashby and Gott (1988). Forty category A and forty category B stimuli from the unidimensional categories were generated by sampling randomly from two bivariate normal distributions (Figure 2.2, left panel). The stimuli for the information integration categories were generated by rotating the 80 rule-based stimuli clockwise by 45º around the center of the spatial frequency-spatial orientation space and then shifting the spatial frequency and spatial orientation by an amount that resulted in the appropriate d' (Figure 2.2, right panel). The category distribution parameters for both structures are listed in Table 2.1.

---

[6] These discriminabilities were chosen to avoid ceiling effects in the UD conditions and floor effects in the II conditions. In addition, we hoped to approximately equate performance across these two category structures in the control condition. To anticipate, we were not successful in equating control condition performance. The II control condition performance was worse than UD control condition performance. However, if the two conditions differed only in difficulty, as sometimes suggested, we would expect a larger dual task interference effect on II than on UD category learning. COVIS, on the other hand, predicts a larger dual task interference effect on UD category learning.

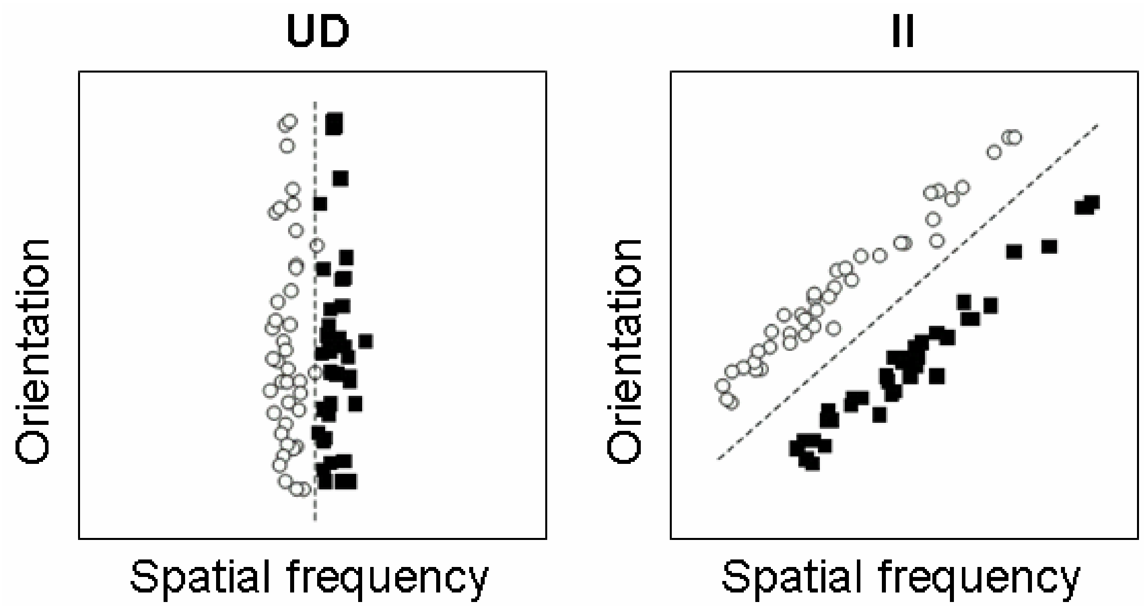

FIGURE 2.2. CATEGORY STRUCTURES USED IN EXPERIMENT 1. UD (left panel): Unidimensional rule-based; II (right panel): information-integration. Open circles denote category A, filled squares denote category B. Dashed lines represent the optimal decision bound.

Category structure	μ_{fA}	μ_{oA}	μ_{fB}	μ_{oB}	σ_f	σ_o	cov_{fo}
Unidimensional rule-based	2.93	45	3.32	45	0.087	34	0
Information-integration	2.81	56	3.44	34	0.674	24	16

TABLE 2.1. CATEGORY DISTRIBUTION PARAMETERS FOR THE CATEGORY STRUCTURES USED IN EXPERIMENT 1. μ = mean; σ = standard deviation; cov = covariance; f = spatial frequency (cycles per degree); o = orientation (degrees); A = category A parameters; B = category B parameters.

Each Gabor patch was generated using Matlab (MathWorks, Natick, MA) routines from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The size of each stimulus was 200 x 200 pixels, covering about four degrees of visual angle and was centered on a computer screen with gray background. Following Waldron and Ashby

(2001), the Stroop task stimuli used in the dual task were two whole numbers sampled without replacement from the range 2-8. On 85% of trials, the numerically larger number was physically smaller (95 pixels tall vs. 180 pixels tall). The stimuli were presented on gray background.

Procedure

Each condition consisted of 5 80-trial blocks of trials. In the control conditions, the participants were told that there were two categories of stimuli and that these are to be learned via corrective feedback. On each trial, a categorization stimulus was presented on the screen and remained there until the participant categorized the stimulus into either category “A” by pressing button “Z” on the keyboard with their left hand or into category “B” by pressing button “?” on the keyboard with their right hand. Corrective feedback was then provided for 1000 ms followed by a 1000 ms delay and 1000 ms inter-trial interval.

In the dual, Stroop conditions, a categorization stimulus was presented centered on the screen, with the Stroop task stimuli presented concurrently to the left and right of the categorization stimulus for 200 ms followed by a rectangular white masks for another 200 ms. The categorization stimulus remained on the screen until the participant categorized it into one of the two categories by pressing “Z” or “?” on the keyboard. The categorization response was followed by 1000 ms corrective feedback and 1000 ms blank screen delay. Then either the word “value” or word “size” appeared on the screen. The participant then indicated on which side the number with the larger value or larger size was presented. The response was followed by 1000 ms corrective feedback and 1000 ms inter-trial interval. The timing of each trial was identical to that used in Waldron and Ashby (2001).

Results

Stroop task performance

Fifty and forty-five participants completed the unidimensional Stroop and information-integration Stroop conditions, respectively. The overall proportion correct for the Stroop task was .84. There was no difference in Stroop task accuracy between the unidimensional (mean = .831, se = .022) and information-integration conditions (mean = .849, se = .019) groups ($t(93) = 0.582$, $p = .562$), suggesting that the effort and cognitive resources allocated to the Stroop task were equal in both groups. Fifteen participants in the UDS condition and thirteen participants is the IIS condition did not reach the 80% required accuracy minimum on the Stroop task and their data were excluded from further analyses.

Category learning performance

For each participant, we computed the proportion correct for each block and the overall proportion correct. We began by examining the shape of the II and UD overall score distributions collapsed across control and Stroop. The distribution of overall scores for the unidimensional category structure deviated significantly from normality (Kolmogorov-Smirnov test (KS) $D(76) = .212$, $p=.002$), while the distribution of overall scores for the information-integration category structure did not (KS $D(66) = .097$, $p = .557$). This pattern held in each block as well. To illustrate, histograms of the overall accuracy distributions for the UDC and IIC conditions are shown in Figure 2.3. While the II distribution is unimodal and close to normal, the UD distribution is bimodal, with one modus close to the chance level of accuracy 0.5 and another at a much higher level of performance.

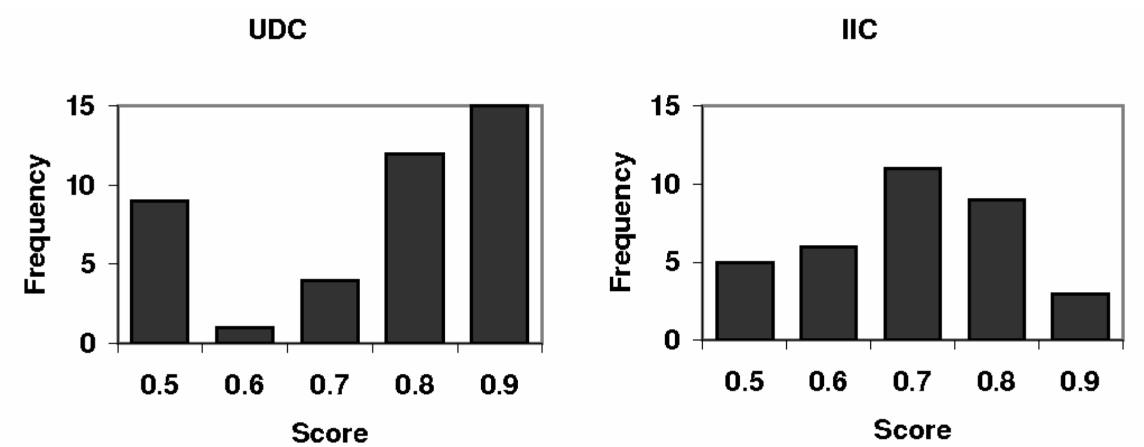

FIGURE 2.3. DISTRIBUTION OF THE OVERALL SCORES IN EXPERIMENT 1. UDC: control unidimensional group; IIC: control information-integration group. Abscissa denotes midpoints of the bins, except the left .5 bin that includes all subjects bellow .55. No subject reached accuracy above .95.

Figure 2.4 presents the mean accuracy scores (proportion correct) for each group. The experimental hypothesis predicts the dual task to have a bigger impact on unidimensional rule-based than on information-integration category learning. To assess the effect of the dual task on underlying distributions with such different shapes provided a challenge as ANOVA, like most standard statistical methods, assumes normal distributions. We used a “bootstrapping”[7] procedure to compare the drop in mean performance across the control and Stroop conditions to determine whether this drop was larger for the unidimensional than for the information-integration category structures. Specifically, the test design was verifying that the 95% confidence interval for the difference in performance drops [(UDC - UDS) – (IIC – IIS)] was reliably bigger than zero. We found that the 15.2% drop in overall performance in the unidimensional rule-based category learning dual task, relative to the control task, was reliably bigger than the

[7] Bootstrap analysis is a statistical method for obtaining an estimate of reliability or error, such as confidence intervals, without a priori assumptions about population distribution. The sample distribution and variability is used as a model for the population distribution and simulations carried out on actual samples are used to draw inference. Bootstrapping is appropriate to use when the distribution shape is unknown (Efron & Tibshirani, 1993).

6.1% drop in the overall performance observed in the information-integration dual task relative to the control task.[8]

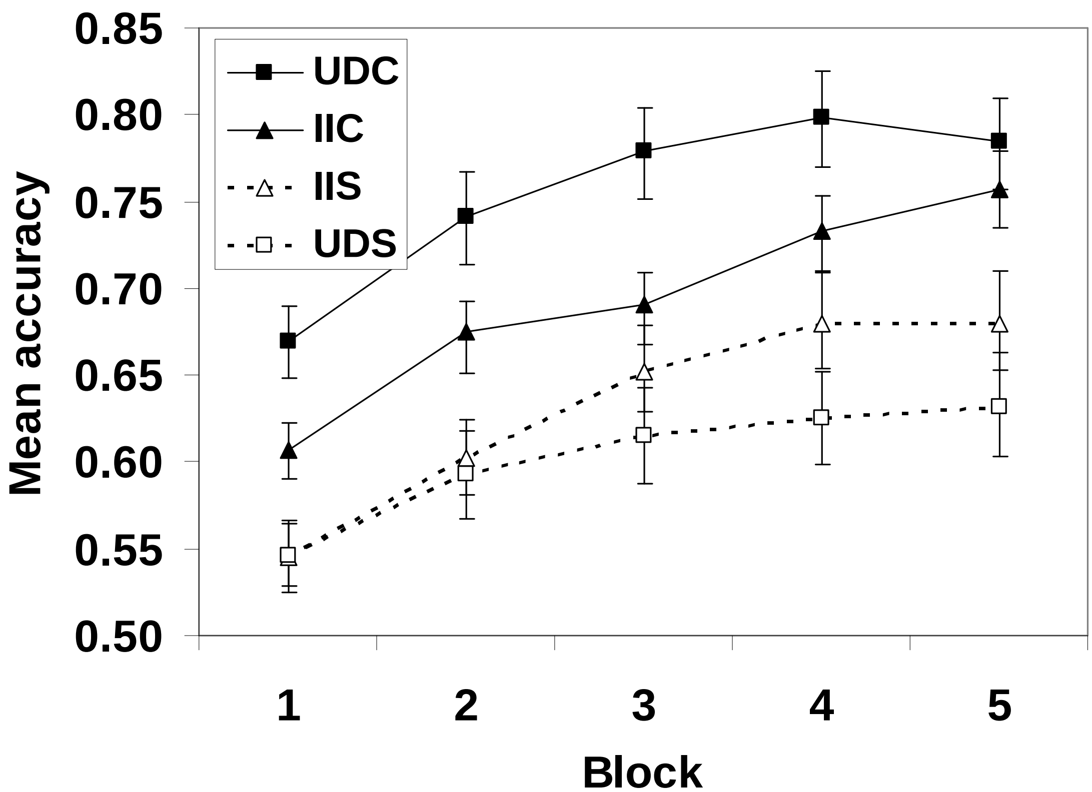

FIGURE 2.4. MEAN CATEGORIZATION BLOCK ACCURACIES (PROPORTION CORRECT) FOR EACH GROUP IN EXPERIMENT 1. The control groups are denoted with solid lines and filled marks, dual Stroop task groups with broken lines and open marks. Unidimensional rule-based groups are marked with squares, information-integration groups with triangles. Error bars denote bootstrapped 68% confidence intervals (equivalent to a standard error of mean).

[8] The difference between the median drop in performance across the two category structures was even stronger. The median performance drop for the unidimensional category observers was 29.0% in the Stroop compared to the control condition which is reliably bigger (bootstrapped 95% confidence interval) than 6.8% median performance drop for the information-integration category observers. For an interested reader, the category structure x condition interaction was detected also using a parametric method (ANOVA interaction [F(1,138) = 4.006, MSE = 0.367, p = .047]).

Discussion

Experiment 1 yielded several interesting results. First, and foremost, including the dual Stoop task had a large effect on unidimensional rule-based, but not information-integration category learning. This finding replicates that observed in Waldron and Ashby (2001) and extends it to a situation in which a large number of normally distributed continuous-valued dimension stimuli were used. Importantly, this pattern holds even though performance is best in the unidimensional control condition and is worst in the unidimensional Stroop condition, ruling out a complexity explanation of the results. Second, the results suggested that unidimensional rule-based category learning (under control and dual task conditions) differed qualitatively from information-integration category learning. Specifically, whereas the distribution of scores observed in the information-integration conditions was unimodal and close to normal, the distribution of scores observed in the rule-based conditions was bimodal, suggesting an all-or-none character to category acquisition (for a similar result, see J. D. Smith, Minda, & Washburn, 2004).

The qualitative difference in the performance profiles across unidimensional rule-based and information-integration conditions is consistent with the multiple-system notion. Rule-based category learning involves the hypothesis-testing system. In this system, different rules are tested and are either accepted or rejected. When the correct categorization rule is identified, categorization accuracy improves dramatically. When incorrect rules are applied, categorization accuracy is near chance, resulting in a bimodal performance distribution. Information-integration category learning involves the procedural learning based system that learns gradually, incrementally and automatically, leading to a normal, unimodal distribution of scores.

EXPERIMENT 2

One potential weakness of Experiment 1 and Waldron and Ashby (2001) is that the number of dimensions relevant for optimal categorization differs across conditions. Nosofsky and Kruschke (2002) pointed out, that the results of Waldron and Ashby are consistent with a single system approach which operates on a single exemplar representation with normal (control) or limited (Stroop) selective attention. To elaborate, Nosofsky and Kruschke (2002) argued that the Stroop task would disrupt ALCOVE's (Kruschke, 1992) selective attention learning parameter. Failure to attend to the single relevant dimension in the unidimensional rule-based task would cause strong interference, because attending to the three irrelevant dimensions would waste vast amounts of processing capacity. In the complex, information-integration category structure, three dimensions are relevant and only one irrelevant. Thus, a wide variety of attentional weights would lead to reasonable performance and only a little processing capacity would be wasted on the one irrelevant dimension.

Ashby and Ell (2002) demonstrated that ALCOVE, although able to account for the qualitative pattern found in Waldron and Ashby (2001), could not account for the quantitative pattern. ALCOVE either underestimates the observed difference between unidimensional control and information-integration control category learning, or it assumes no attention learning in the Stroop condition (leading participants in the unidimensional Stroop condition to be unaware that a single dimension was relevant).

In Experiment 2, we decided to investigate the notion of Nosofsky and Krushke's using a conjunctive rule-based category structure where both dimensions are relevant for optimal categorization (see Method section for details). Nosofsky and Kruschke would predict none or small dual task interference because no dimension is irrelevant in this task and a wide range of attentional weights can provide a high level of performance. Also, because the attention learning mechanism is disrupted, attention should be spread over both dimensions throughout the course of learning. COVIS, however, would predict stronger dual task interference because the conjunctive task, unlike the information-integration task, is solved under control condition by the hypothesis testing system. Under the dual Stroop task condition, use of conjunctive rules (attending to both dimensions) is less likely and use of sub-optimal unidimensional rules (selective attention to one dimension while ignoring the other) is more likely, because conjunctive rules require more working memory capacity than unidimensional rules. Experiment 2 aims to provide a test of these two alternatives.

Method

Participants

Sixty students at The University of Texas at Austin participated in the experiment in partial fulfillment of a class requirement or for pay. Thirty completed the conjunctive control (CJC) condition and thirty completed the conjunctive Stroop (CJS) condition. All participants were tested for 20/20 vision.

Stimuli and apparatus

The stimuli, stimulus generation procedure and apparatus were identical to those used in Experiment 1. The only difference was in the nature of the category structures. Eighty stimuli were generated by sampling randomly from four bivariate normal distributions. Three were assigned to category A and one to category B. The four distributions parameters and the number of stimuli generated from each are displayed in Table 2.2. A scatter plot of the stimuli and the optimal rule is presented in Figure 2.5. The optimal rule required participants to respond B when the spatial frequency was high and the orientation was steep, and to respond A otherwise. Note that both dimensions are relevant for correct categorization. The number of stimuli generated from each distribution was chosen to equate the number of stimuli in both categories and in an attempt to reduce the usage of unidimensional rules to solve the task.

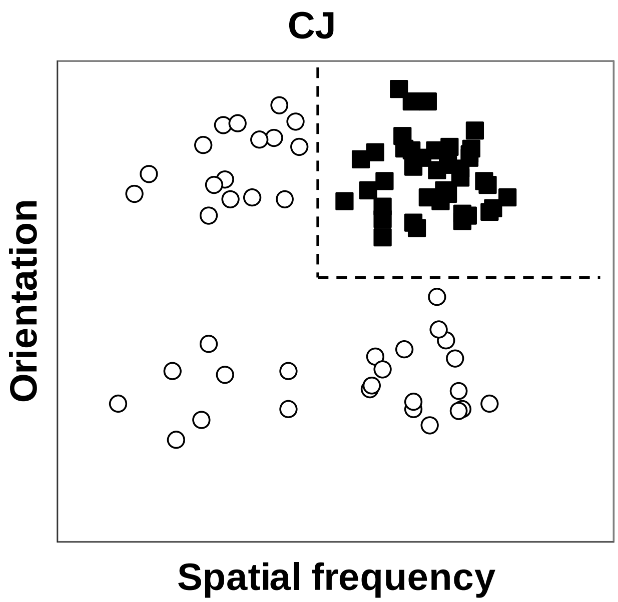

FIGURE 2.5. CONJUNCTIVE (CJ) CATEGORY STRUCTURE USED IN EXPERIMENT 2. Open circles denote category A, filled squares denote category B. Dashed line represents the optimal decision bound.

Distribution	μ_f	μ_o	σ_f	σ_o	cov_{xy}	N
A_1	2.95	35	0.21	3.12	0	8
A_2	3.30	35	0.21	3.12	0	16
A_3	2.95	55	0.21	3.12	0	16
B	3.30	55	0.21	3.12	0	40

TABLE 2.2. DISTRIBUTION PARAMETERS FOR THE CONJUNCTIVE CATEGORY STRUCTURE USED IN EXPERIMENT 2. N: number of stimuli derived from each distribution. Stimuli from the distributions A_1, A_2, A_3 were all members of category A.

Procedure

The procedure was identical to that in Experiment 1 except that there were four rather than five blocks of 80 trials. Participants were told that perfect performance is possible and that they should certainly achieve above 80% correct before the end of training.

Results

Stroop task performance.

Mean Stroop task accuracy was .862 (sem = .026). Five participants did not reach required 80% Stroop task accuracy, and their data were excluded from further analyses.

Categorization task performance

We first inspected the distribution of scores in order to compare them to those in Experiment 1 (data not shown). Although there was a tendency toward bimodality, the Kolmogorov-Smirnov test did not show a significant deviation from normality for either condition nor collapsed across the control and the dual condition in any block (KS D(55) = .153, p = .153 for the collapsed data and overall score distribution).

Mean categorization accuracy for each block of trials is shown in Figure 2.6. Overall categorization accuracy was 70.2% in the control group and 60.6% in the Stroop group. Thus, the Stroop task produced a 9.6% drop in categorization accuracy that was significant (bootstrapped 95% confidence interval for the drop (CJC – CJS)).

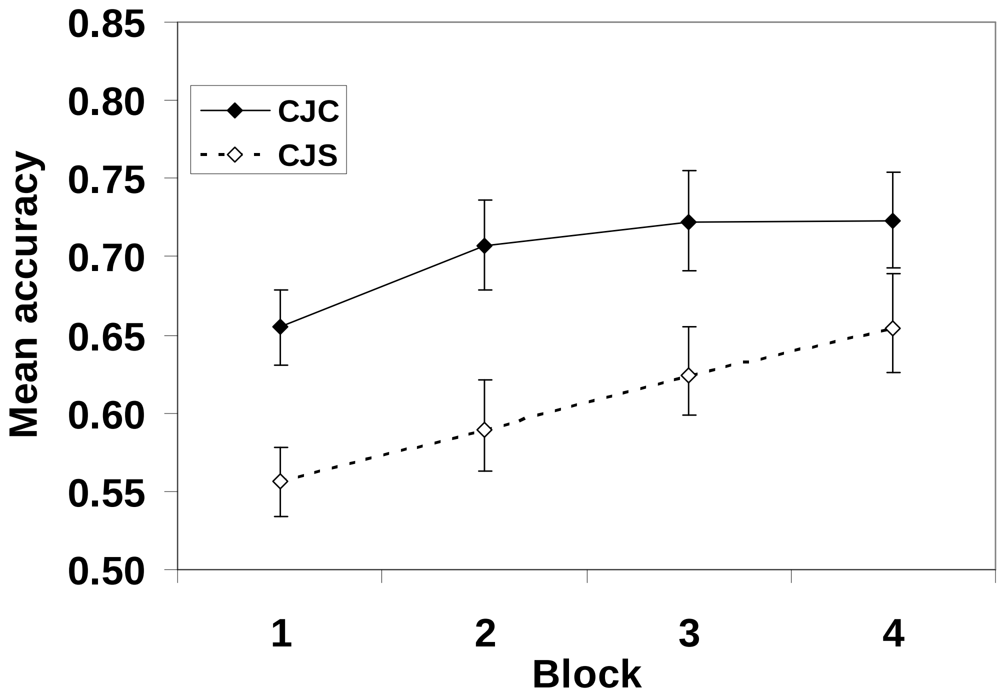

FIGURE 2.6. MEAN BLOCK ACCURACIES IN EXPERIMENT 2 (CONJUNCTIVE TASK). Control group is denoted with solid line and filled diamonds, dual Stroop task group with broken line and open diamonds. Error bars denotes bootstrapped 68% confidence intervals (equivalent to a standard error of mean).

To examine response strategies, we fit a general linear classifier, a conjunctive and a unidimensional decision bound model to each participant's responses in the last block (the details of these models can be found e.g. in Maddox, Ashby, & Bohil, 2003). We found that the proportion of participants who used a strategy employing both dimensions dropped from 77% in the control to 44% in the dual task condition and the proportion of participants using a unidimensional rule for categorization increased from 7% in control to 17% in the dual task condition.[9]

[9] The responses of the rest of the participants were best account by a flat response strategy that assumes that the response is independent from the stimulus values.

Discussion

The results from Experiment 2 supported the COVIS prediction that categorization based on the combination of both dimensions is less likely and using unidimensional strategies more likely under dual task condition than under control condition; a prediction that is opposite of that from Nosofsky and Kruschke's (2002) account of Waldron and Ashby's (2001) results. To compare the observed drop in performance in the conjunctive rule-based condition with those from Experiment 1, we computed the average performance across the first four blocks of trials (since Experiment 2 concluded four rather than five blocks of trials) in each condition. The results are displayed in Figure 2.7. Figure 2.7 suggests that the impact of the dual Stroop task was indeed larger on conjunctive rule-based than on information-integration category learning, as predicted by COVIS, and opposite from that predicted by Nosofsky and Kruschke (2002).

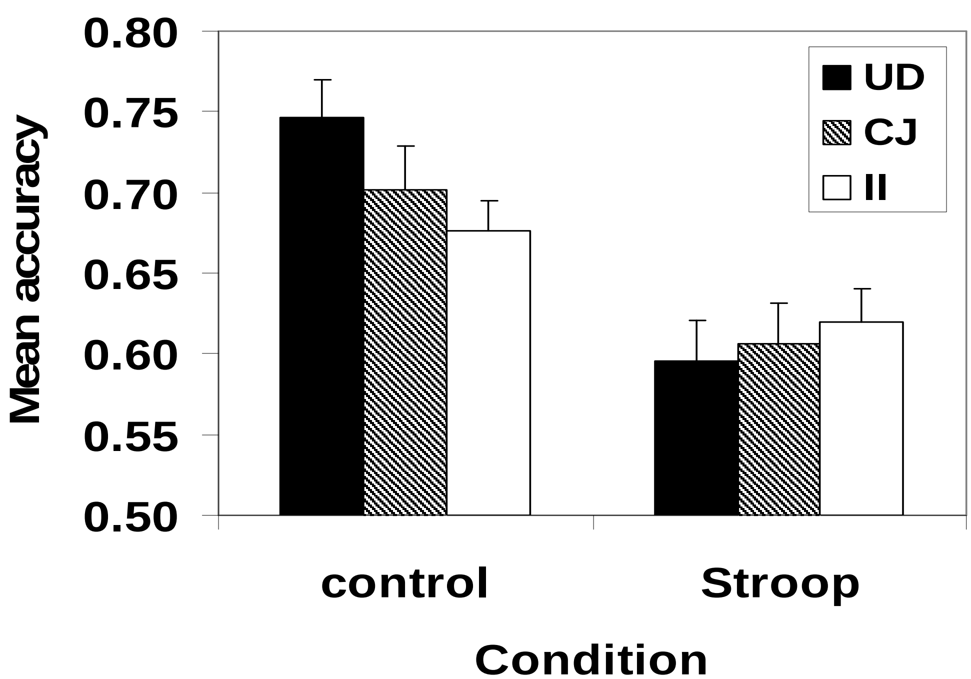

FIGURE 2.7. COMPARISON OF MEAN CATEGORIZATION ACCURACIES ACROSS EXPERIMENTS 1 AND 2. Only data from the first four blocks were used.

Both the shape of the score distributions and the dual task effect on the conjunctive task were intermediate between those found in Experiment 1 for unidimensional rule-based and information-integration. The results suggest that a number of strategies may be used to resolve the conjunctive task and each of these strategies may be influenced by the dual task differently. Detailed discussion of the dual task interference for the three category structures is reserved for the General Discussion.

GENERAL DISSCUSSION

The theoretical framework which gave rise to the experiments reported in this article was the COVIS model of category learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). COVIS builds upon a body of research that identified alternative strategies of category learning and extends it by identifying the underlying neural structures. This line of research contrasts with theories assuming a single system of category learning. In this discussion, we will first focus on the COVIS account of the observed pattern of data, then review alternative multiple-system approaches to categorization, and finally ask whether a single system approach to categorization may be sufficient in accounting for the results observed here and elsewhere.

COVIS

COVIS assumes the existence of at least two category learning systems: an initially favored hypothesis-testing system that seeks explicit rules and relies on working memory and selective attention, and an implicit system that is procedural–learning-based and essentially automatic. Two predictions result from this notion. First, category learning by the hypothesis testing system when a simple correct categorization rule exists that yields nearly perfect performance (such as our unidimensional rule-based category structure) should have an all-or-none character, while learning by the procedural-learning based system is gradual and incremental. Second, a dual task requiring limited cognitive resources, working memory and selective attention, should impair the hypothesis testing system, but not the procedural system.

The three category structures used in the two studies reported in this paper differ in their level of attainability by the two systems. Unidimensional rule-based category learning resulted in a bimodal, all-or-none distribution of scores and was affected most by the dual task, suggesting a strong reliance on the hypothesis-testing system in solving the task. The unidimensional category structure is indeed well acquired by the hypothesis-testing system, because a simple rule can yield almost perfect accuracy. However, if the correct rule is not found, alternative rules yield performance at chance levels of accuracy. The procedural learning based system may exhibit poor acquisition of such a structure because the variance along the relevant dimension is small while the variance along the irrelevant dimension is high. The high convergence of connections from the inferotemporal cortex to the tail of caudate nucleus may cause the same striatal units to be activated by stimuli coming from different categories but sharing similar value on the irrelevant dimension, making the stimulus-response mapping within the caudate difficult. Thus, although this task was easiest under the control condition, the need to find the one correct rule by the hypothesis-testing system with limited resources and unreliable responses from the implicit system made it most difficult under the dual task condition.

Neuroimaging study by Bench et al. (1993) showed that the anterior cingulate and frontal cortex are structures strongly activated while performing a Stroop task. The fact that the presence of the Stroop task most affected unidimensional rule-based category learning provides an empirical support for the COVIS proposition that the explicit, hypothesis testing system, but not the procedural system, relies on working memory and attentional processes and on these same underlying brain structures (i.e. anterior cingulate and frontal cortex). The dual Stroop task may influence several stages of the hypothesis-testing system. It may make selective attention to relevant dimension more difficult to achieve because selective attention is needed for the Stroop task. Its working memory load may make it harder to remember the current rule to be tested and which rules did not work previously. It may impair the ability to detect conflict and evaluate performance and select and switch to a new rule (anterior cingulate functions).

Information-integration category learning was the most difficult in the control condition, but exhibited the smallest decrement in performance in the dual task condition, becoming the easiest. The information-integration category structure is better acquired by the procedural system than the hypothesis-testing system. COVIS predicts that after trying unsuccessfully all salient rules, the weight on the hypothesis-testing system decreases and the responses become dominated more often by the procedural system. The procedural system learns the stimulus-response mapping gradually and incrementally, yielding a normal distribution of scores. The stimulus-response mapping in the caudate is facilitated by the larger distance of the stimuli from the two categories in the stimulus space (d prime in the physical space of 10.3 compared to 4.5 for the unidimensional structure). The dual Stroop task may influence information-integration category learning in two opposite ways: It may hurt performance because it reduces cognitive resources needed for initially biased hypothesis testing and/or slows down the shift in favor of the implicit system, or it may facilitate performance because the limited capacity hypothesis-testing system becomes less initially biased and/or the overall system shifts faster towards the implicit system. We found a slight performance drop in the dual compared to the control condition, suggesting that the first influence or a combination of both is more likely. Because the implicit system itself is unaffected by the dual task, once the weight of that system increases sufficiently, accuracy would be expected to be essentially the same under the control as under the dual condition.

The conjunctive rule-based category structure used in the Experiment 2 has properties intermediate between those used in Experiment 1, yielding intermediate difficulty and performance drop under the two conditions. The optimal conjunctive rule yields the highest accuracy (100% possible), however, unidimensional rules on either dimension can provide accuracy up to 80%, and information-integration strategies may be successful as well due to relative high separability of the four underlying distributions. The proportion of participants using a combination of both dimensions (spreading attentional weights in terms of ALCOVE (Kruschke, 1992)) for categorization decision decreased and the proportion of participants using values on a single dimension increased under the dual task condition, contrary to the ALCOVE prediction (Nosofsky & Kruschke, 2002) and with agreement to the prediction assuming two learning systems. The results of Experiment 2 also argue for dual Stroop task interference on performance evaluation and rule switching, in addition to working memory load, because participants were more likely to stick with the suboptimal unidimensional rules despite corrective feedback.

Rule versus similarity

There has been a long tradition in cognitive psychology research of focusing on the distinction between perceptual categorization that is based on a rule application and that based on overall similarity to previously seen instances (Allen & Brooks, 1991; Brooks, 1978; Folstein & Van Petten, 2004; Kemler-Nelson, 1984; J. D. Smith & Shapiro, 1989). Rule versus similarity distinction provides an alternative theory of multiple strategies of categorization. Rule application involves a high working memory load and requires analytic, serial processing of criterial attributes with differential weighting of attributes, while similarity-based processing involves a low working memory load and holistic, parallel, automatic processing with equal weighting of attributes (E. E. Smith, Patalano, & Jonides, 1998).

The theories assuming alternative strategies of categorization involving qualitatively distinct processes of rule application and similarity judgment are strikingly similar to the computational level description of the COVIS model. The rule application is assumed to involve working memory and selective attention to criterial attributes, similar to the explicit, hypothesis testing system. The similarity judgment is an automatic, holistic process that does not have a high working memory load, comparable to the implicit, procedural learning based system. A dual task reduces the likelihood of using analytical rules in categorization (J. D. Smith & Shapiro, 1989).

Rule versus similarity distinction theories would therefore predict a similar pattern of results as obtained here, because the holistic strategies promoted over the analytic strategies under the dual task have different relative utilities for correct categorization. However, direct application of these theories to the results from our two experiments is complicated by the dissimilarity of the experimental paradigms. Experiments illustrating the dissociation between rules and similarity often use a unitary category structure where category membership can be determined perfectly from rule application or similarity-based processes and induction of either process is achieved by instruction manipulation (e.g. explicit formulation of the rule versus feedback training only in Allen and Brooks, 1991). An alternative are experiments using real world categories for which the existence or nonexistence of necessary attributes (rules) is known to participants (e.g. size of a quarter in Rips, 1989). In our experiments, participants have no prior knowledge about the nature of the category structure and training is based on feedback only for all category structures. Category structures themselves, rather than instruction or prior knowledge, promote or inhibit the use of either system. The process of rule discovery and testing is of equal importance to rule application in the COVIS model, and the interaction and relative weighting of the two systems is explicitly stated. On the other hand, studies in which use of a rule versus a similarity judgment is addressed in conditions in which both strategies are available at any given trial may help to shed more light on how the competition between the hypothesis testing and the implicit system is resolved. Also, while the rule versus similarity distinction may be widely valid across modalities and extend to higher level cognition, such as language, COVIS has a narrower focus on visual perceptual categorization and, because of the specified underlying neurobiology, cannot be automatically applied outside its original domain. In sum, despite some methodological and terminological differences, the neuropsychological COVIS model and the cognitive psychology based theories of alternative rule and similarity strategies of categorization are more likely to complement than to oppose one another.

Single versus multiple systems in category learning

The alternative to the notion of multiple systems in categorization is the notion of a single categorization system. First we have to make clear what *a system* means. In COVIS, the two category learning systems operating in parallel differ both in the computational and implementation level of description; one system coding for explicit rules in frontal structures using selective attention and working memory, the other encoding instances in the inferotemporal cortex and procedural learning-based stimulus-response mapping in the striatum. Both systems then compete (or cooperate) to determine the response of the overall system (the organism). When arguing against the multiple system account of Waldron and Ashby (2001) results, Nosofsky and Kruschke (2002) accept that other processes, such as selective attention to relevant dimension, may take place in category learning. However, they emphasize that different processes operate on a single exemplar category representation. What seems like a distinction between categorization based on a rule application versus overall similarity evaluation is then the same exemplar-based categorization when all attention weight is on one diagnostic dimension versus when attention weight is spread about equally across many dimensions. A similar idea was recently presented by Pothos (2005) who argues that rules and similarity represent two extremes on a single continuum of similarity operations, with no need to model rule and similarity processes separately. Rule application is a similarity evaluation process where only a single or small number of object's features are involved.

These are compelling ideas and the imperative of parsimony requires accepting the single system notion (a single representation with a single process) unless we have a sufficient body of evidence that a single system explanation cannot account for the empirical data available.

Several lines of evidence lead us to believe that a single system explanation is not sufficient to account for the data observed in our two experiments and, most importantly, in the complex of a broad range of other studies. First, we already discussed how Nosofsky and Kruschke (2002) account of Waldron and Ashby (2001) is inconsistent with the results of our Experiment 2 (see also Ashby and Ell, 2002, for evaluation of Nosofsky and Kruschke).

Second, single system models, such as ALCOVE (Kruschke, 1992) do not specifically address the underlying neural substrate for the exemplar storage and the neural mechanism of categorization. A number of studies focused on differential activation of the brain in different categorization paradigms and suggest that at least humans have available more than one system involving different neural circuits and category representations (E. E. Smith, Patalano & Jonides, 1998; Seger & Cincotta, 2002; see Keri, 2003, for a review of studies including clinical neuropsychology findings, functional neuroimaging, and single cell studies). To address specifically the single versus multiple representations issue, Keri (2003) summarizes a number of studies showing that inferotemporal cortex is responsible for category instances representation, while prefrontal cortex encodes abstract rules. Such a finding supports at least two-fold category representation, one is the representation of specific instances (exemplars), the other of rules. Because an organism behaves as an integral system, different representations and processes will interact and act in concordance in order to produce meaningful behavior. Pothos' (2005) putative continuum from rules to similarity may thus reflect relative contribution of each system to the overall response of an organism, such as postulated in COVIS, by the relative weighing of the two subsystems' responses in producing the final decision.

Third, even if a single system model could account for the pattern of data observed in our two experiments, we may still question whether that provides us with a valid and more parsimonious explanation. First, exemplar models are often viewed as highly flexible. Recently, Olsson, Wennerholm and Lyxzen (2004) showed that exemplar and other mathematical models often suffer from overfitting, i.e. accounting perfectly for noise as well as actual variance due to cognitive processes. Second, with respect to the issue of parsimony, it is unclear whether a single system model that requires different sets of assumptions (and parameter values) about the underlying processes to account for the wide array of "multiple systems" data is more parsimonious than a multiple systems model that *a priori* predicts the patterns observed in the multiple systems data. For instance, Nosofsky and Kruschke (2002) claim: “As long as the sensitivity parameter *c* is not too high, ALCOVE predicts far greater interference on the simple one-dimensional task than on the complex three-dimensional task” (p.171). The *c* parameter measures the overall discriminability of the stimuli and *should* be high for such highly discriminable stimuli as used in Waldron and Ashby (2001). We conclude that the dual task interference effects found here provide evidence for the existence of at least two category learning systems, one that utilizes working memory and attention and dominates in learning of the rule-based category structures and one that does not require these limited resources and dominates in learning of the information-integration category structures.

Chapter 3: Working Memory in Category Learning[10]

WORKING MEMORY ROLE IN THE HYPOTHESIS TESTING AND PROCEDURAL LEARNING SYSTEM

Most multiple systems models of category learning postulate a hypothesis-testing (or rule-based) component. These include models that make no claims about the neurobiological underpinnings, such as RULEX (Nosofsky, Palmeri, & McKinley, 1994) or ATRIUM (Erickson & Kruschke, 1998), as well as neurobiologically-inspired models like COVIS (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; see also Patalano, Smith, Jonides, & Koeppe, 2001). COVIS postulates two category learning systems: a hypothesis-testing system that seeks verbalizable rules and is mediated primarily by the prefrontal cortex, and a procedural learning system that learns stimulus-response associations primarily mediated by striatum. As stated explicitly in COVIS (and perhaps implicitly assumed in RULEX and ATRIUM), processing of the feedback in the hypothesis testing system is effortful and requires working memory and attentional capacity. Following feedback on an incorrect trial, a number of events occur in the hypothesis testing system: (1) The salience of the current rule decreases; (2) A decision is made about whether to re-use the current rule or to generate and select a new one; (3) If applicable, attention is switched from the old rule to the new rule. These events in the hypothesis testing system require both time and availability of the limited resources (working memory and attention: Maddox, Ashby, Ing, & Pickering, 2004; Waldron & Ashby, 2001). On the other hand, the procedural system requires the feedback to follow the categorization response immediately (Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005), but feedback is then processed automatically and does not require working memory or attention (Maddox, Ashby, Ing, & Pickering, 2004).

[10] Major portions of this chapter have been previously published as an article Zeithamova & Maddox (2007). The role of visuospatial and verbal working memory in perceptual category learning. Memory and Cognition, 35(6), 1380-1398. Copyright 2007 Psychonomic Society, Inc

Maddox et al. (2004) tested the prediction that feedback processing is effortful and time consuming for the hypothesis-testing system, but not for the procedural system. They contrasted rule-based and information-integration category learning using the category structures depicted in Figure 2.2 (previous chapter) across the three experimental conditions displayed in Figure 3.1.A. In the control condition, participants viewed a stimulus, generated a categorization response and received 500 ms of corrective feedback followed by a 2000 ms (blank screen) inter-trial interval. In the "long" feedback processing condition, the categorization response feedback was followed by 2500 ms blank screen delay display to allow feedback processing, after which a trial of a Sternberg's *verbal* working memory task was presented. Trial was concluded with a 2000 ms inter-trial interval. As the verbal working memory task, the Sternberg (1966) memory scanning task was used (Figure 3.1.B). In the Sternberg memory scanning task, four digits between 1 and 9 were presented for 500 ms (memory set). Next, a 1000 ms delay (blank screen) was presented followed by a single digit (probe). The participants had to indicate whether this digit was a part of the memory set or not. In the "short" condition, the categorization response feedback was followed immediately by the working memory task.

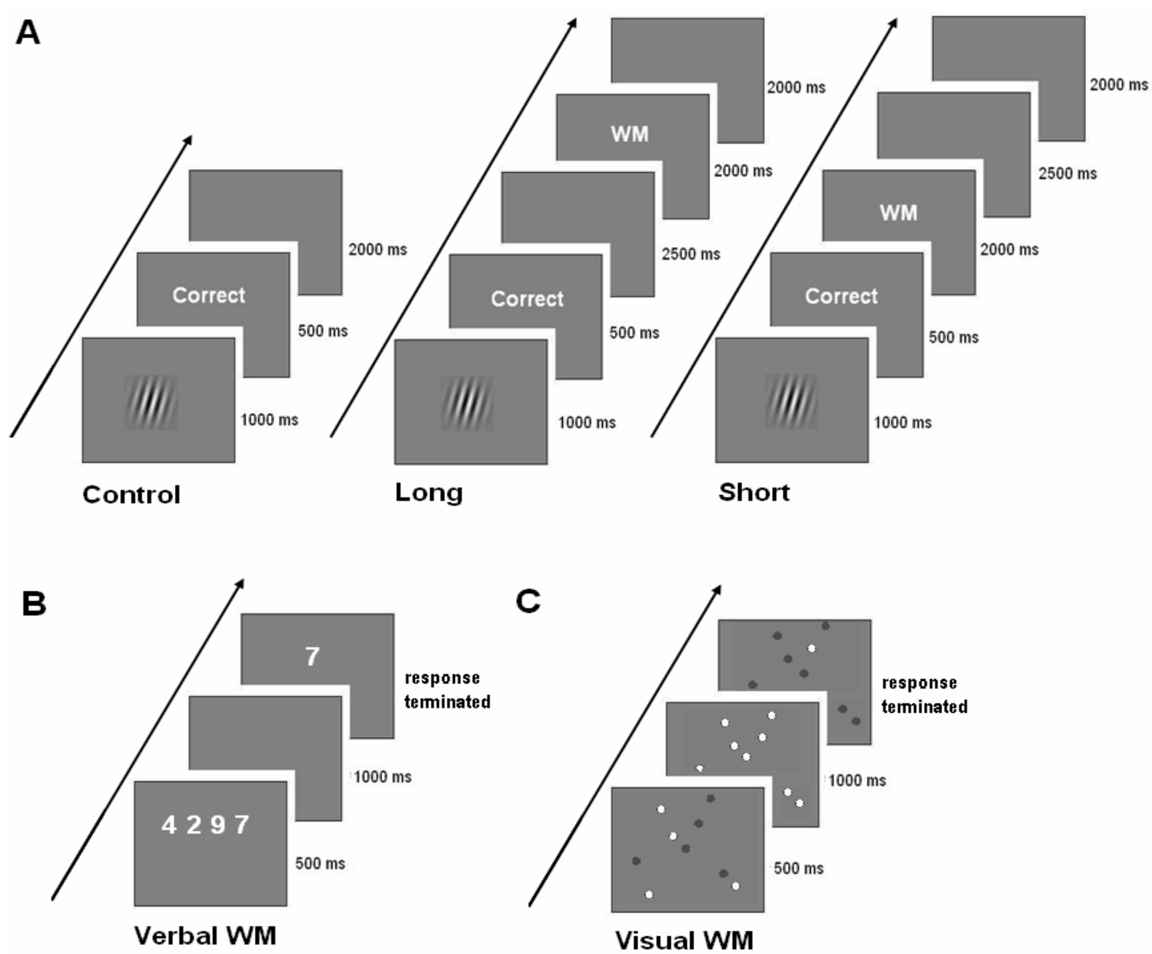

FIGURE 3.1. TRIAL DESIGN IN MADDOX ET AL. (2004) AND EXPERIMENT 1. (A) Schema of three conditions: Control = control condition, long = long feedback processing time condition, short = short feedback processing time condition; WM = working memory task. (B) Verbal working memory task design. (C) Visuo-spatial working memory task design.

Maddox et al. found a large disruption of rule-based category learning in the "short" condition compared to the control condition, and a small (and not statistically significant) disruption of rule-based category learning in the "long" condition. Presence of the Sternberg task had no effect on information-integration category learning.

OVERVIEW OF THE CURRENT STUDIES

Maddox et al. (2004) provided evidence that working memory and attention are necessary for accurate feedback processing in rule-based category learning, but not information-integration category learning. They used a version of Sternberg's memory scanning task to tax working memory and attentional processes. However, several questions remain to be answered.

In the working memory literature, a distinction is made between at least two kinds of working memory: verbal and visuo-spatial (e.g. Baddeley & Logie, 1999; Jonides et al., 1996). This distinction holds behaviorally (Cocchini, Logie, Sala, MacPherson, & Baddeley, 2002; Logie, Zucco, & Baddeley, 1990; Shah & Miyake, 1996), and there is also evidence that verbal and visuo-spatial working memory rely on different neural systems (Goldman-Rakic, 1998; E. E. Smith, Jonides, & Koeppe, 1996). However, the existing models of the hypothesis testing system do not address the distinction and do not make any a priori prediction regarding the effects of a secondary verbal vs. visuo-spatial working memory task on category learning. The Sternberg task is a standard *verbal* working memory task (Raghavachari et al., 2001), the question remains how a *visuo-spatial* working memory task affects rule-based and information-integration category learning. In this section, we discuss a series of reasonable predictions.

Maddox et al. (2004) showed that a sequential verbal working memory task did not affect information-integration category learning. One hypothesis is that this result would be replicated with a visuo-spatial working memory task. The logic is as follows: if the information-integration task is primarily learned via the procedural learning system and the procedural system processes feedback automatically without the need for (any kind of) working memory or attentional resources, there should be no effect of visuo-spatial working memory on information-integration category learning. A second hypothesis is that the presence of a visuo-spatial workin memory task will affect information-integration category learning. The logic is as follows: processing in the procedural system depends critically on the visual stimulus representation in the inferotemporal cortex (Freedman, Riesenhuber, Poggio, & Miller, 2003). This representation may be disrupted by the presence of a visuo-spatial task because the stimuli of the visuo-spatial task are visually presented and difficult to encode verbally. Additionally, the procedural system is assumed to rely on basal ganglia (e.g. Filoteo, Maddox, Salmon, & Song, 2005; Poldrack, 2002), much like visuo-spatial working memory is assumed to depend on the basal ganglia, perhaps even to a larger extend than other types of working memory (Lawrence, Watkins, Sahakian, Hodges, & Robbins, 2000; Postle, Jonides, Smith, & Corkin, 1997).

In the hypothesis-testing system, working memory is needed for holding the currently active rule, comparing the rule with the current feedback, and selecting and switching to a new rule if necessary. Because rules learned by the hypothesis-testing system are usually verbalizable (Ashby, Alfonso-Reese, Turken, & Waldron, 1998), one reasonable prediction is that a sequential verbal working memory task will adversely affect rule-based category learning, but a sequential visuo-spatial working memory task will not. This prediction assumes that rule-based learning involves generating a verbal representation of the stimulus, response and feedback. Thus, placing a load on a separate visuo-spatial working memory store will not affect these verbal processes. Although verbal and visuo-spatial working memory stores are separable, a second possibility is that a sequential visuo-spatial working memory task will adversely affect rule-based category learning. Two general mechanisms may underlie such a disruption. First, a visuo-spatial working memory task may affect rule-based learning indirectly, because it, like the verbal working memory task, relies on the central executive as a common, limited-capacity resource (Baddeley, 1995; Baddeley & Logie, 1999). A load on the visuo-spatial working memory store would thus influence rule-based category learning via the central executive or general attention demand. Second, visuo-spatial working memory may be involved in some aspect of rule-based category learning that does not require verbal working memory.

To conceive which aspects of rule-based learning may be differentially influenced by the secondary verbal and visuo-spatial working memory tasks, we need to consider what steps or processes may possibly be parts of rule learning and discovery. With unidimensional rules on stimuli that vary along a limited number of continuously varied dimensions (as used in our experiments), the process may include the following intertwined steps: (1) selection and focused attention on one *stimulus dimension* (e.g. “spatial frequency”); (2) generation, representation and testing of a *categorization rule* (in the narrow sense of the meaning) along that dimension (e.g. “narrow stripes are category A, wide stripes are category B”); (3) learning, storing and application of a *categorization criterion* (e.g. the optimal spatial frequency distinguishing between narrow and wide stripes). Verbal working memory would then seem critical primarily for rule generation, maintenance and testing (Step 2), while visuo-spatial working memory may then be critical for learning and representation of the actual categorization criterion (image of a particular spatial frequency: Step 3) and/or for identification of individual stimulus dimensions (analytical decomposition of the stimulus) that is a basis for Step 1. A second reasonable prediction is thus that the visuo-spatial working memory may disrupt rule-based learning, either in a similar fashing to a verbal working memory task, or differently.

The goal of Experiment 1 was to test these hypotheses using the same procedure as Maddox et al. (2004), but replacing the verbal Sternberg working memory task with a visuo-spatial working memory analog. The basic experimental design is depicted in Figure 3.1.A, the design of the verbal Sternberg task and our visuo-spatial analog are presented in Figure 3.1.B and 3.1.C. To anticipate, we found no effect of the sequential visuo-spatial working memory task on information-integration category learning, but we did find an effect on rule-based category learning.

Experiment 2 explored the generality of the working memory effects in perceptual category learning. Whereas Maddox et al. (2004) and Experiment 1 examined a rule on the spatial frequency of a Gabor patch stimulus, Experiment 2 examined rules on the orientation of a Gabor patch stimulus. Gabor patch stimuli have several desirable properties for perceptual category learning researchers. For example, they have a known dimensional structure with two separable dimensions. The two dimensions have simple verbal labels, are measured in different units and have no emergent properties. However, orientation has two special properties that spatial frequency does not have. First, it is periodic with zero degrees being equivalent to 360 degrees. Second, it contains special values, called "cardinal" orientations. Cardinal orientations – vertical and horizontal– are processed differently from other values at both the neural and behavioral level. People are more sensitive and more accurate when asked to judge orientations around cardinal orientations (Campbell & Kulikowski, 1966; Heeley & Timney, 1988; Orban, Vandenbussche, & Vogels, 1984). This is likely due to the fact that more neurons in primary visual cortex are tuned to cardinal orientations (Furmanski & Engel, 2000).

In Experiment 2, we examined the effect of both visuo-spatial and verbal working memory tasks on rule-based learning when the optimal categorization rule required the participant to separate orientations above and below 70 degrees from horizontal (oblique – Experiment 2A), or orientations above and below 90 degrees (cardinal – Experiment 2B). When the criterion is set at 70 degrees, we expect the findings to replicate those observed when spatial frequency was relevant, because learning a rule on the (arbitrary) orientation of 70 degrees likely requires the same processes needed to learn a rule on spatial frequency. In both cases, there are two obvious dimensions (spatial frequency and orientation) along which to generate explicit rules, these rules need to be tested, the irrelevant dimension ignored and the optimal criterion on the relevant dimension needs to be learned. We will test this prediction in Experiment 2A.

The processes involved in cardinal orientation based category learning may be however different. As we noted above, people exhibit greater sensitivity to orientation changes near cardinal orientations (Campbell & Kulikowski, 1966; Heeley & Timney, 1988; Orban, Vandenbussche, & Vogels, 1984). We say that cardinal orientations are *perceptually special*. This well established higher perceptual sensitivity may likely lead to more precise categorization (leading to higher asymptotic accuracy for learners) once a correct rule is discovered. In other words, perceptual advantage for cardinal orientations should improve *categorization criterion* learning.

Furthermore, it is possible that cardinal orientations do not simply constitute an easier *categorization criterion* value on the general *categorization rule* "respond A if orientation is greater than a criterion, respond B if orientation is smaller than a criterion." Rather, as soon as the participant notices that stimuli vary in orientation, the cardinal orientations may constitute salient, spontaneously used categorization rules that create intuitive categories (e.g. right-tilted, left-tilted). We say that cardinal orientations may be *conceptually special*. In other words, cardinal orientations may constitute a highly salient *categorization rule* per se, with Step 2 and Step 3 being merged together. To our knowledge, no categorization studies explicitly tested this assumption. However, Huttenlocher, Hedges and Duncan (1991) reported that their participants used cardinal orientations as reference points in location estimation. If cardinal orientations are conceptually special, people may tend to use these early in learning as a rule of the first choice and most of the categorization rule discovery stage of learning may be skipped. This possible conceptual significance of cardinal orientations may lead to higher proportion of learners and much more rapid learning with minimal working memory load (and thus minimal effect of a secondary working memory task) as most participants would simply select the correct categorization rule as their first choice.

Now let us consider possible effects of a working memory task when learning a rule on a cardinal orientation. First consider verbal working memory effects. If cardinal orientations are special perceptually, but not conceptually, all the effortful hypothesis-testing processes still need to take place to find the correct rule, and we would expect to observe a verbal working memory task effect. If cardinal orientations are special perceptually *and* conceptually, the highly salient cardinal rule will be chosen early and much of the hypothesis testing process will be bypassed. Under these conditions, we would expect to see no or a minimal effect of the verbal working memory task.

Second, consider visuo-spatial working memory effects. If the mechanism of the visuo-spatial working memory task effect is the same as that for the verbal working memory task (e.g. through the central executive), then we predict the same effect (or lack of effect) for both types of working memory tasks. If the mechanisms of visuo-spatial and verbal working memory task effects differ, then the predictions will differ depending on the role of the visuo-spatial working memory in rule-based category learning.

Let us return to the three steps that take a part in unidimensional rule-based learning: (1) selection and focused attention on one stimulus dimension, (2) categorization rule generation, representation and testing, and (3) criterion learning, representation and application. In order for the first step to occur, the participant needs to notice how the stimuli vary between one another and decompose them into their individual constituent dimensions. If visuo-spatial working memory is necessary for such analytic perception of individual stimulus dimensions, the participant may have difficulty identifying the dimensions along which the stimulus varies and how they vary, interfering with the first step of rule-based learning and leading to a learning deficit in the visual condition (even if the verbal working memory task had no effect). If visuo-spatial working memory is crucial for learning and representing of the categorization criterion (Step 3), a cardinal orientation criterion may lead to two opposing scenarios. First, learning of a criterion on the cardinal orientation may lead to little or no visuo-spatial working memory effect. This may follow either from (a) the existing higher perceptual sensitivity around cardinal orientation which should make learning of cardinal orientation criterion easier and less working memory demanding or (b) from possible conceptual significance of the cardinal orientation, i.e. if sorting based on a cardinal orientation criterion constitutes an intuitive, highly salient categorization rule per se and does not need to be actually learned the same way as an oblique criterion. Second, and making an opposite prediction, a criterion on a cardinal orientation may be more working memory demanding because the increased perceptual sensitivity to orientations around cardinal would lead participants to consider and test larger number of possible criteria. This argument assumes that cardinal orientations are not conceptually special and a criterion on a cardinal orientation needs to be learned in very much the same way as on an oblique orientation. In other words, it assumes that a participant would equally likely consider a criterion e.g. on 88 degrees as a criterion on 90 degrees. We will test the working memory effects on cardinal orientation based categorization in Experiment 2B.

EXPERIMENT 1

The goal of Experiment 1 was to test a number of hypotheses regarding the effect of a visuo-spatial working memory task on rule-based and information-integration category learning. To achieve this goal, we replicated Maddox et al.'s (2004) procedure, but replaced the Sternberg (verbal) working memory task with a visuo-spatial working memory task (Figure 3.1). We were interested in determining whether the effects on rule-based and information-integration category learning observed for a verbal working memory task replicated when it was replaced with a visuo-spatial working memory task. If the pattern was not replicated, we wanted to determine how the pattern changed.

Method

Participants and Design

Two hundred ninety three students at The University of Texas at Austin participated in the experiment in partial fulfillment of a class requirement or for pay. All participants were tested for 20/20 vision. The experimental design was 2 category structures (rule-based vs. information-integration) x 3 sequential working memory conditions (control, long feedback processing time, short feedback processing time). Each participant completed one of the 6 experimental conditions: rule-based control (RB control: 46 participants), rule-based long feedback processing time (RB long: 51 participants), rule-based short feedback processing time (RB short: 53 participants), information-integration control (II control: 38 participants), information-integration long feedback processing time (II long: 52 participants), or information-integration short feedback processing time (II short: 53 participants).

Stimuli and Apparatus

Category learing. The stimulus dimensions and category structures were identical to those from Maddox et al. (2004). The categorization stimuli were Gabor patches (sinewave gratings enclosed in a Gaussian envelope) that varied across trials in spatial frequency and orientation. For the rule-based and information-integration category structures, forty category A and forty category B stimuli were obtained by randomly sampling from two bivariate normal distributions. The rule-based task was unidimensional, with spatial frequency being the relevant dimension and orientation being the irrelevant dimension. The optimal rule was to "respond A if the frequency of the Gabor is small (below 3.13 cycles per degree), respond B if the frequency of the Gabor is large (above 3.13 cycles per degree)." Both dimensions were relevant in the information-integration task and no simple verbal rule discriminated between the two categories. A schematic representation of the two category structures is depicted in Figure 2.2 (Chapter 2); the category distribution parameters for both category structures are listed in Table 3.1.

Category structure	μ_{fA}	μ_{oA}	μ_{fB}	μ_{oB}	σ_f	σ_o	cov_{fo}
Rule-based	2.97	45	3.28	45	0.087	34	0
Information-integration	2.84	55	3.41	35	0.674	24	16

TABLE 3.1. CATEGORY DISTRIBUTION PARAMETERS FOR THE RULE-BASE AND INFORMATION-INTEGRATION CATEGORY STRUCTURES USED IN EXPERIMENT 1. μ = mean; σ = standard deviation; cov = covariance; f = spatial frequency (cycles per degree); o = orientation (degrees); A = category A parameters; B = category B parameters.

Each Gabor stimulus was generated and presented using Matlab (MathWorks, Natick, MA) running Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The stimuli were 200 x 200 pixels, centered on a computer screen, covering about four degrees of visual angle.

Visuo-spatial working memory. A visuo-spatial working memory task that was analogous to the Sternberg working memory task used in Maddox et al. (2004) was created (Figure 3.1.C). The participant was asked to remember 4 locations out of 9 possible (analogous to remembering four numerical digits sampled from 9 possible digits). First, nine locations were randomly placed in an imaginary 9 x 9 grid, with the restriction that there is one location in each imaginary row and one location in each imaginary column. The nine locations were marked by dark gray circles each with a radius of 48 pixels and remained visible throughout the visuo-spatial task trial. After 500 ms, four out of the nine locations were highlighted by a white circle with radius of 40 pixels for 500 ms (memory set). The participant needed to remember those four locations. Next, all nine locations were highlighted for 1000 ms (delay period). Finally, only one location (probe) was highlighted and the participant's task was to indicate whether that location was one of the four initially highlighted locations. The probability of a probe being one of the memory set was .5.

Procedure

The procedure was identical to that of Maddox et al. (2004). There were three conditions: control, long and short (Figure 3.1.A). Each condition consisted of four randomly ordered 80-trial blocks. The participants were informed that there were two equally likely categories and that their task is to learn which patterns go into which category via corrective feedback. In the control condition, a categorization stimulus was presented on each trial and remained on the screen for 1000 ms or until the participant categorized it in either category A or category B. If the participant did not respond during the 1000 ms, the Gabor stimulus disappeared and only the response prompt ("Categorize the pattern as A or B") remained on the screen. The participant had as much time as needed to make a response. Corrective feedback was then provided for 500 ms followed by a 2000 ms inter-trial interval (blank screen).

In the long feedback processing time condition, the categorization response feedback was followed by a 2500 ms blank screen to allow feedback processing, after which a trial of visuo-spatial working memory task was presented. The visuo-spatial task response was followed by a 2000 ms blank screen inter-trial interval and no feedback was provided. The short feedback processing time condition was similar to the long feedback processing time condition, however, the categorization response feedback was followed immediately by the visuo-spatial working memory task and the 2500 ms delay was placed after the working memory task response. After each block of 80 trials, participants were given a short self-paced break during which they were informed how many trials had passed and were urged to keep their visuo-spatial task accuracy high.

Results and Discussion

Working memory task performance

The mean percent correct in the working memory task was high, 96.0% for RB long (sd = 7.0%); 94.1% for RB short (sd = 7.6%); 96.0% for II long (sd = 7%); 93.2% for II short (sd = 9.8%) groups. There were no differences between rule-based (mean = 95.0%, sd = 7.3%) and information-integration (mean = 94.5%, sd = 8.6%) category structure groups ($t(207) = 0.413$, $p = .680$), suggesting that the resources allocated to the working memory task were distributed about equally. There was a difference between participants in long (mean = 96.0%, sd = 7.0%) and short (mean = 93.6%, sd = 8.7%) feedback processing time conditions collapsed over the two category structures ($t(207) = 2.125$, $p = .035$).

Category learning performance

Distribution of accuracy scores. In Chapter 2, we identified substantial differences between the distributions of accuracy scores in rule-based and information-integration category learning. We found a bimodal distribution of scores in rule-based learning and a normal distribution of scores in information-integration learning. Thus, we began our analysis of category learning performance by examining the distribution of accuracy scores. The distribution of scores for the rule-based groups and information-integration groups (collapsed over the three working memory conditions) from the final block of trials are presented in Figure 3.2. As is apparent from Figure 3.2, the score distribution for the rule-based category structure participants deviates from normality (Kolmogorov-Smirnov D(150) = .156, p = .001), while the score distribution for the information-integration category structure participants does not deviate from normality (KS D(143) = .056, p = .510). The same pattern holds for each of the working memory groups (control, long and short) within each category structure, with the rule-based groups appearing bimodal and the information-integration groups appearing normal. These results suggest that different processes underlie rule-based and information-integration category learning that lead to very different performance profiles.

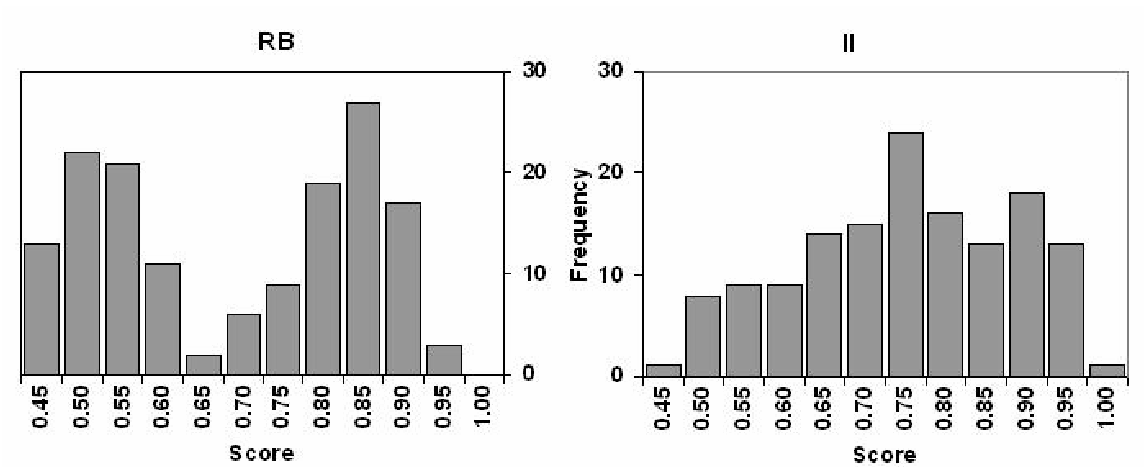

FIGURE 3.2. DISTRIBUTION OF THE FINAL BLOCK ACCURACY IN EXPERIMENT 1. Scores of rule-based (RB) and information-integration (II) category learning groups collapsed over the three feedback processing conditions (control, short and long).

To gain further insight into the nature of the bimodality in the rule-based score distributions, we applied a series of models separately to the control, long and short score distributions. Each model assumes that the observed score distribution results from a mixture of two underlying distributions. A detailed description of the models and the results are presented in Supplemental Data. To summarize, the score distribution for each of the three rule-based category learning groups was best fit by a mixture of two underlying distributions, one with mean at .5 (chance) and one with a mean around .8. As suggested by this analysis, there were two types of participants in each condition: learners, who discovered the appropriate categorization rule, and chance performers, who failed to discover the rule. The difference between the three rule-based groups was in the relative weight of the two distributions, i.e. the proportion of participants that fell under the .5 modus (nonlearners) and under the .8 modus (learners). More specifically, the effect of the secondary visuo-spatial working memory task was to decrease the proportion of learners (participants who discovered the rule and constituted the .8 accuracy modus) from 69.2% in the RB control group to 44.0% in the RB long and 48.8% in the RB short group.

Proportion of learners. The distributional analyses suggest that the score distributions in the rule-based conditions are composed of a mixture of two populations of participants (chance performers and learners), and that the relative ratio of learners to chance performers decreases when a sequential working memory task is included. As a further test, we compared the proportion of learners in each condition by defining “learners” as participants who reached .65 correct and higher in the last block of trials and “nonlearners” as participants who failed to reach .65 correct in the last block of trials. We choose .65 correct as it appears to be a natural cut-off in the score distributions (see Figure 3.3, left panel) and constitutes an average between the means of the chance distribution (.5) and the high performance distribution (.8) for the rule-based groups. The results are depicted in Figure 3.3 (left panel).

In the RB control group, 32 out of 46 participants learned. There were significantly fewer learners in both the RB long (23 out of 51, $\chi(1)=5.897$, $p=.015$) and RB short conditions (26 out of 53, $\chi(1)=4.26$, $p=.039$), compared to the RB control condition. In the II control group, 30 out of 38 participants learned. The proportions of learners in II long (36 of 52) and II short (36 of 53) conditions were not significantly different from that observed in the II control condition [long vs. control: $\chi(1)=1.060$, $p=.303$; short vs. control: $\chi(1)=1.350$, $p=.245$]. It is worth noting that these proportions of learner analyses for the information-integration groups were included for completeness, and to compare with the analyses for the rule-based conditions. This learning criterion is less meaningful for the information-integration conditions as it cuts the distribution of scores at an arbitrary value.

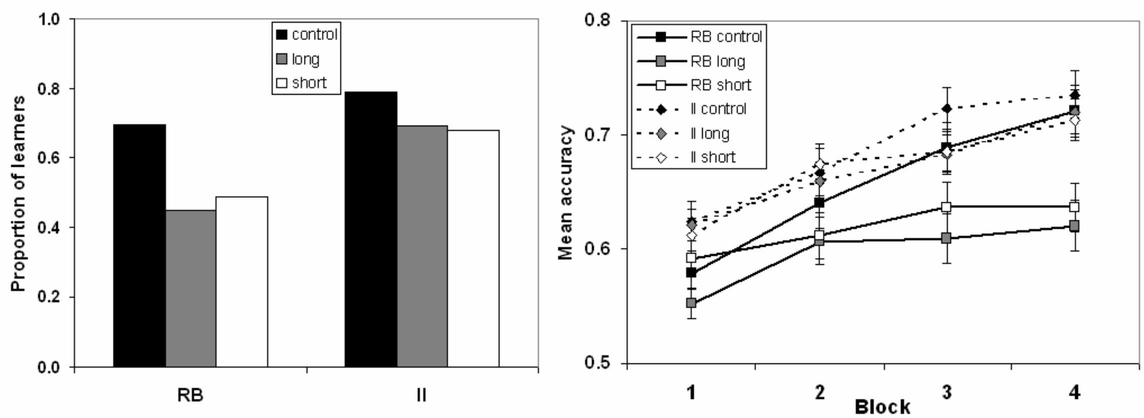

FIGURE 3.3. RESULTS OF EXPERIMENT 1. Left panel: Proportion of learners in Experiment 1. RB = rule-based groups, II = information-integration groups. Right panel: Mean accuracy (proportion correct) for each group in Experiment 1. Unidimensional rule-based (RB) groups are denoted with square symbols and solid lines, information-integration (II) groups with diamond symbols and broken lines. Error bars denote bootstrapped 68% confidence intervals (equivalent to a standard error of mean).

Mean proportion correct. Unlike proportion of learners, mean accuracy (proportion of correct responses), is a more suitable performance measure for the information-integration groups. For the rule-based groups, mean accuracy is also applicable, but reflects the relative proportion of two populations of participants rather than reflecting performance of a typical participant. We decided to use bootstrapping procedures (Efron & Tibshirani, 1993) that are more appropriate than traditional parametric statistics when the distribution shape is non-normal or unknown. Categorization accuracy for each group in each block of trials is presented in Figure 3.3 (right panel).

We found a significant effect of the secondary visuo-spatial task on rule-based category learning but not on information-integration category learning. Average final block performance in the RB long group dropped by .10 relative to the RB control group (bootstrapped p(control = long) = .002). Performance in the RB short group dropped by .084 relative to the RB control (p(control = short) = .009).[11]

There were no significant differences among the three information-integration category learning groups in any block of trials or in overall accuracy. Average performance in the final block from the II control group was only .015 higher than that observed in the long group (bootstrapped p(control=long) = .602), and was only .022 higher than that observed in the II short group (p(control=short) = .415).

Brief summary

To summarize the results, we found an adverse effect of the visuo-spatial working memory task on rule-based, but not information-integration category learning. The lack of interference between the visuo-spatial working memory and the information-integration category learning task is interesting, as both tasks use visual stimuli and are thought to rely on basal ganglia (visuo-spatial working memory: e.g. Lawrence, Sahakian, Hodges, & Rosser, 1996; Lawrence, Watkins, Sahakian, Hodges, & Robbins, 2000; discrimination and category learning: e.g. Packard & McGaugh, 1992; Poldrack, 2002). This result, together with that observed in Maddox et al. (2004) with a verbal working memory task, provides support for the assumption that learning of the information-integration category structure is mediated by a procedural system which processes feedback automatically, without relying on attention or working memory.

The significant effect of visuo-spatial working memory on rule-based category learning replicates the effect observed in Maddox et al. with a verbal working memory task, and extends it to a visuo-spatial working memory task. The effect of the verbal working memory task on rule-based learning found in Maddox et al. (2004) was expected, because attention and (verbal) working memory have been implicated in rule generation, rule maintenance, rule selection and rule switching (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Dougherty & Hunter, 2003). On the surface, none of these

[11] The tested null hypothesis is mean1 – mean2 = 0. All bootstrapped p-values for the difference in means reflects a conservative, two-tailed hypothesis and may be converted to one-tailed values when divided by 2.

processes appears to require visuo-spatial working memory, yet visuo-spatial working memory affected rule-based category learning.

As outlined in the introduction, visuo-spatial working memory task may affect rule-based learning indirectly via cognitive resources shared with verbal working memory or may have a direct effect through mechanisms other than those impacted by the verbal working memory task. We will discuss this issue in more detail in the General Discussion. Let us now turn to Experiment 2 that examines the generalizability of the working memory effects in rule-based learning and may add evidence in favor of one or the other explanations of the visuo-spatial working memory effects observed in Experiment 1.

EXPERIMENT 2A

In Experiment 2A, we tested the effects of a sequentially presented visuo-spatial or verbal working memory task on rule-based learning when the optimal categorization rule was to "respond A when the orientation of the stimulus is larger than 70 degrees from horizontal, respond B when the orientation of the stimulus is smaller than 70 degrees from horizontal." We compared a control condition with a short feedback processing visuo-spatial working memory condition (as in Experiment 1), and with a short feedback processing verbal working memory condition (as in Maddox et al, 2004). We dropped the long feedback processing condition because its effects were modest in Maddox et al. (2004).

Method

Participants and Design

Seventy-two students at The University of Texas at Austin participated in the experiment in partial fulfillment of a class requirement or for pay. All participants were tested for 20/20 vision. Each participant completed one of 3 experimental conditions: (1) control (no secondary task); (2) visual (each category learning trial was immediately followed by a visuo-spatial working memory task trial); or (3) verbal (each category learning trial was immediately followed by a Sternberg verbal working memory trial). There were 24 participants in each condition.

Stimuli and Apparatus

Category learning task. The stimuli were Gabor patches that varied across trials in spatial frequency and orientation. A rule-based category structure was used with the optimal rule being "respond A if the orientation of the Gabor is larger than 70 degrees, respond B if the orientation of the Gabor is smaller than 70 degrees." The stimuli were randomly sampled from two bivariate normal distributions both with a mean spatial frequency of 3.13 cycles per degree and a standard deviation of 0.95 cycles per degree. Category A stimuli had a mean orientation of 75.8º with a standard deviation of 3.1º. Category B stimuli had a mean orientation of 64.2º with a standard deviation of 3.1º. Orientation and spatial frequency of the Gabors were uncorrelated. Forty category A and forty category B stimuli were generated. A schematic representation of the category structure is depicted in Figure 3.4, left panel. The apparatus was identical to that from Experiment 1.

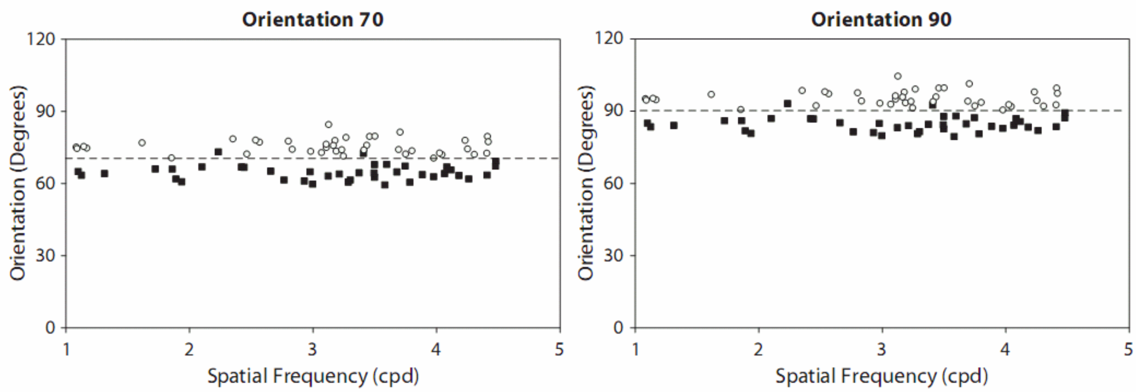

FIGURE 3.4. CATEGORY STRUCTURES USED IN EXPERIMENT 2. Left: Experiment 2A (Orientation 70); Right: Experiment 2B (Orientation 90). Open circles denote category A stimuli, filled squares denote category B stimuli. cpd = cycles per degree

Working memory tasks. The visuo-spatial task was identical to that used in Experiment 1 (Figure 3.1.C). The verbal working memory task (Figure 3.1.B) was taken from Maddox et al. (2004). On each trial, four digits (memory set) were randomly selected without replacement from the set of digits from 1 to 9. The memory set was displayed for 500 ms in a horizontal array centered on the screen and spanning approximately 8 degrees of visual angle horizontally and 4 degrees of visual angle vertically. A blank screen followed for 1000 ms. Finally, a single digit (probe) was presented in the center of the screen and the participant was asked to indicate whether the probe was a part of the memory set. The probability that the probe was a member of the memory set was .5.

Procedure

The procedure was similar to that used in Experiment 1 with the exception that we equated the duration of the inter-trial interval across conditions and added a fixation cross to prepare participants for the next trial. Each condition consisted of four blocks of eighty randomly ordered trials. The participants were informed that there were two equally likely categories and that their task was to learn which pattern goes into which category via corrective feedback. In the control condition, each trial started with a 500 ms fixation cross (a plus sign) to prepare the participant for the upcoming trial. A categorization stimulus was then presented and remained on the screen for 1000 ms or until the participant categorized it in either category A or category B. Corrective feedback was provided for 500 ms followed by 2500 ms inter-trial interval (blank screen).

The two working memory (visual and verbal) conditions were similar to the "short" condition from Experiment 1. Each trial also started with a 500 ms fixation cross, followed by a categorization stimulus presented for 1000 ms or until the participant responded. Corrective feedback was presented for 500 ms after a response was made and a working memory task trial immediately followed (verbal, Figure 3.1.B or visuo-spatial, Figure 3.1.C). After the participant responded, a 2500 ms inter-trial interval (blank screen) concluded the trial, and no working memory task feedback was provided. As in Experiment 1, after each block of 80 trials, participants were given a short self-paced break during which they were informed how many trials had passed and were urged to keep their working memory task accuracy high.

Results

Working memory task performance

The mean percent correct was 95.5% in the visuo-spatial task (sd = 4.5%) and 96.6% in the verbal task (sd = 3.4%). This difference was not statistically significant ($t(46) = .940$, $p = .352$).

Category learning performance

<u>Distribution of scores and proportion of learners.</u> As expected from Experiment 1, the distribution of scores in the final block of trials violated normality (Kolmogorov Smirnov $D(72)=.221$, $p=.002$), and instead was bimodal (Figure 3.5, left panel). Using the same procedure applied in Experiment 1 (Supplemental Data), we found that the

score distribution for each of the three conditions was best fit by a mixture of two underlying distributions, one with mean at .5 (chance) and one with mean around .9. The effect of both secondary working memory tasks was again in decrease of the proportion of participants who discovered the rule (learners) and constituted the .9 accuracy modus, from 83.3% in the control group to 50.0% in the visual and 50.0% in the verbal condition. Compared to Experiment 1, both the mean performance level for the learners and relative proportion of learners in the observed score distribution was higher. Specifically, whereas the mean performance level for learners in Experiment 1 was about .8, it was .9 in Experiment 2A. Similarly, whereas the relative proportion of learners was 69.6% in Experiment 1 control condition, it was 83.3% in Experiment 2A control condition. Taken together, these data suggest that the rule on the orientation of the Gabor stimuli was somewhat easier to learn.

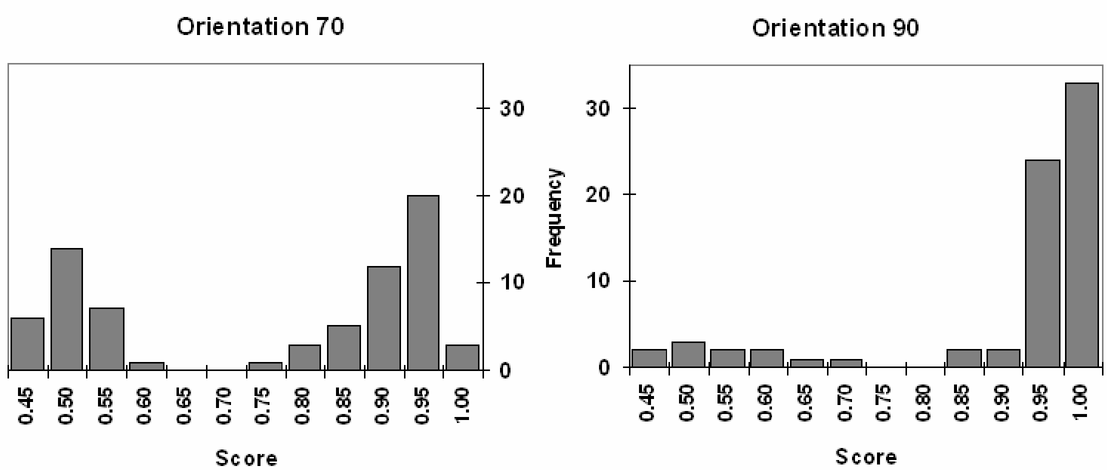

FIGURE 3.5. DISTRIBUTION OF THE FINAL BLOCK ACCURACY SCORES IN EXPERIMENT 2. Orientation 70: Experiment 2A; Orientation 90: Experiment 2B. Scores collapsed over the three working memory conditions (control, visual, verbal).

We also analyzed the proportion of learners in each group using the same criterion as in Experiment 1 (at least .65 correct in the last block). The results are depicted in Figure 3.6 (left panel). There were 20 learners (out of 24) in the control condition, which is significantly more than 12 learners (out of 24) in either the visual or the verbal condition ($\chi^2(1) = 6.0$, p=.014).[12]

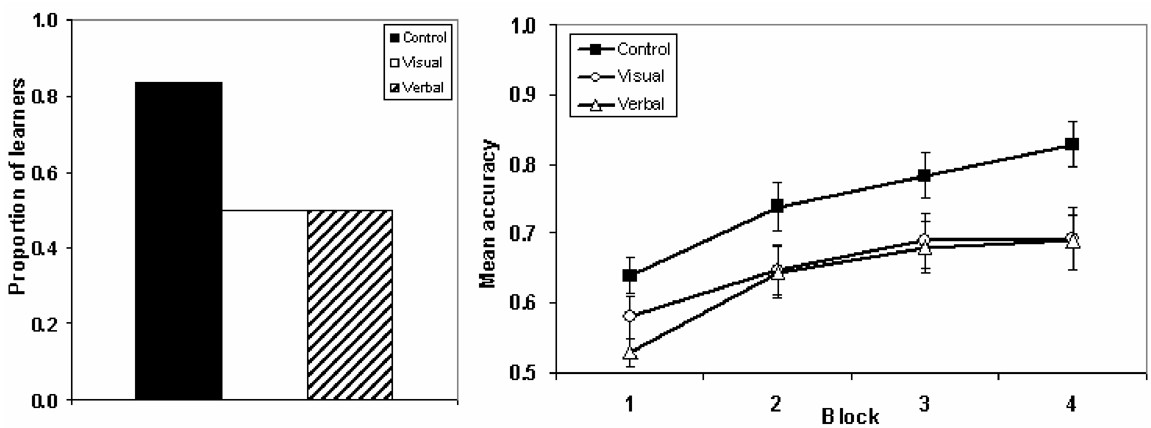

FIGURE 3.6. RESULTS OF EXPERIMENT 2A. Left panel: Proportion of learners. Right panel: Mean accuracy in each block. Error bars denote bootstrapped 68% confidence intervals (equivalent to a standard error of mean).

Mean proportion correct. Block accuracies in Experiment 2A are presented in Figure3.6 (right panel). We found that both the visuo-spatial and verbal working memory tasks significantly disrupted category learning. Specifically, the average final block performance dropped by .134 in the visual condition relative to the control condition (bootstrapped p(control = visual) = .017), and by .139 in the verbal condition relative to the control condition (bootstrapped p(control=verbal) = .009).

To summarize, we found an adverse effect of both visuo-spatial and verbal working memory tasks on rule-based category learning when the criterion was on an (arbitrary) oblique orientation. As expected, these results are similar to those found when the criterion was on spatial frequency of a Gabor, suggesting that these working memory

[12] The learner criterion of .7, which is the average of the .5 chance score distribution and .9 high performing score distribution in Experiment 2A, gives identical results, as no subject had final block accuracy between .65 and .70 (see Figure 6, left panel).

effects generalize. We now turn to Experiment 2B that tests more fully the generality of these effects.

EXPERIMENT 2B

In Experiment 2B, we used a category structure that was formally identical to that in Experiment 2A, except that the criterion was shifted from 70 degrees to 90 degrees, a cardinal orientation (Figure 3.4, right panel). As stated in the Introduction to Experiment 2, we suspect that this slight manipulation may have a substantial effect on rule-based category learning under both the control condition and when a sequentially presented working memory task is included. The two main aims of Experiment 2B were to investigate whether cardinal orientations are conceptually special and whether we find evidence for dissociability of the visuo-spatial and verbal working memory effects.

Method

Participants and Design

Seventy-two students at The University of Texas at Austin participated in the experiment in partial fulfillment of a class requirement or for pay. All participants were tested for 20/20 vision. Each participant completed one of 3 experimental conditions: (1) control, (2) visual, or (3) verbal. There were 24 participants assigned into each condition. No student participated in both Experiment 2A and Experiment 2B.

Stimuli, Apparatus and Procedure

<u>Category learning task</u>. The stimuli, apparatus and category structure were identical to that from Experiment 2A except that the optimal categorization rule was to “respond A if the orientation of the Gabor is larger than 90 degrees (is left-tilted), respond B if the orientation of the Gabor is smaller than 90 degrees (is right-tilted).” The

stimuli were randomly sampled from two bivariate normal distributions that both had a mean spatial frequency of 6.25 cycles per degree with a standard deviation of 2.14 cycles per degree. Category A stimuli had a mean orientation of 95.8º with a standard deviation of 3.1º. Category B stimuli had a mean orientation of 84.2º with a standard deviation of 3.1º. The orientation and spatial frequency of the Gabors were uncorrelated. Forty category A and forty category B stimuli were generated. A schematic representation of the category structure is depicted in Figure 5 (right panel). The secondary working memory tasks and experimental procedure were identical to those from Experiment 2A.

Results

Working memory task performance

The mean percent correct was 95.8% in the visuo-spatial task (sd = 3.2%) and was 96.6% for the verbal task (sd = 2.9%). This difference was not statistically significant ($t(46) = .898$, $p = .374$). These accuracies are very similar to those obtained in Experiment 2A.

Category learning performance

<u>Distribution of scores, proportion of learners and mean proportion correct.</u> The distribution of scores in the final block of trials deviated from normality (Kolmogorov-Smirnov $D(72)=.373$, $p<.001$), but unlike in Experiments 1 and 2A, it did not appear bimodal (Figure 3.5, right panel). Rather, the vast majority of participants learned the correct categorization rule and performed at a high rate of accuracy (above .90 correct). In addition, there were no differences among groups in the proportion of learners (Figure 3.7, left panel). Only 2 participants in the control condition, 3 participants in the visual condition and 4 participants in the verbal condition did not reach .65 proportion correct in the last block of trials. Block accuracies in Experiment 2B are presented in Figure 3.7 (right panel). As is apparent from the figure, there were small and nonsignificant differences among the groups during the final block of trials (bootstrapped

p(control=visual)=.270, p(control=verbal)=.232) and performance was high. There were performance differences in the initial block of trials. In the first block, performance in the visual condition dropped by .148 compared to the control condition (p(control=visual)=.004) and performance in the verbal condition dropped by .108 compared to the control condition (p(control=verbal)=.008).

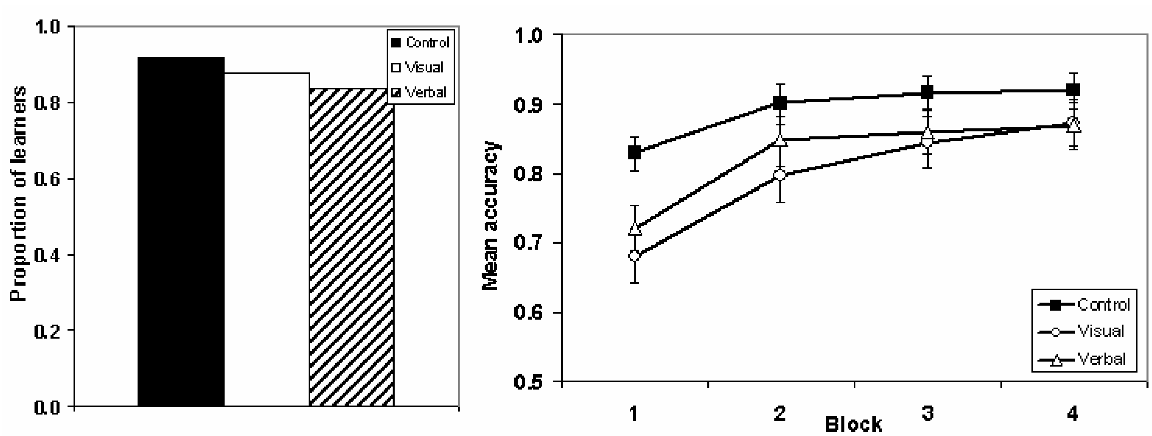

FIGURE 3.7. RESULTS OF EXPERIMENT 2B. Left panel: Proportion of learners. Right panel: Mean accuracy in each block. Error bars denote bootstrapped 68% confidence intervals (equivalent to a standard error of mean).

A comparison of the control conditions from Experiment 2A and Experiment 2B suggests that cardinal orientations are special both perceptually *and* conceptually. The higher asymptotic accuracy for learners in Experiment 2B compared to Experiment 2A (e.g. compare left and right panels in Figure 3.5) supports the notion that cardinal orientations are perceptually special. A comparison of early control performance in Experiment 2A and Experiment 2B (e.g. compare Block 1 in Figure 3.6, right panel with Block 1 in Figure 3.7, right panel) supports the notion that cardinal orientations are also conceptually special and constitute a highly salient categorization rule. High accuracy in the control condition starting at Block 1 suggests that a majority of the participants selected the correct categorization rule very early on.

In Experiments 1 and 2A, the secondary working memory task led primarily to a decrease of the proportion of learners compared to the control condition. This pattern did not replicate in Experiment 2B because the proportion of learners was high and approximately equal in all conditions by the end of training. Because there were no differences in the proportion of learners among the conditions and almost every participant learned the rule, we turned to an alternative measure of performance that may be more sensitive to detect an effect of a secondary task, if there was any. We turned to a measure often used in categorization research – the number of trials needed to reach an accuracy criterion (*trials to criterion*). As the vast majority of participants in Experiment 2B succeeded in learning the task, using trials to criterion is suitable to characterize the speed of learning in all conditions. The trials to criterion measure is not suitable in Experiment 1 and Experiment 2A, as half or more of the participants in the visual and verbal working memory conditiontions failed to reach any reasonable learning criterion, whereas the majority of the participants in the control conditions did.

<u>Trials to criterion</u>. As one trials-to-criterion measure, we recorded for each participant the trial on which the accuracy over the last 80 trials first reached or exceeded .65 correct (the learning criterion used in the previous experiments). This analysis is presented in Figure 3.8 (C6580; i.e. criterion of .65 correct over previous 80 trials). Participants in the visuo-spatial working memory condition needed significantly more trials to reach the criterion (bootstrapped p(control=visual)=.013) than participants in the control condition, whereas there was no difference in the trials-to-criterion between the verbal working memory group and the control group (bootstrapped p(control=verbal)=.859). Because a large proportion of participants reached the .65 correct criterion right at the trial number 80 (i.e, discovered the rule during the first 80 trials of the experiment), we examined a number of other performance criteria and smaller window sizes to ensure that the results were robust. We typically found a large and significant effect of the visuo-spatial working memory task and a small and non-significant effect of the verbal working memory task relative to the control condition. For example, for the criterion of .75 correct over the last 40 trials (Figure 3.8, C7540), we found that the visuo-spatial working memory group needed on average 48 more trials than the control group to reach the criterion (bootstrapped p(control=visual) = .002),

while the verbal working memory group needed on average only 12 more trials than the control group (bootstrapped p(control=verbal) = .127).[13]

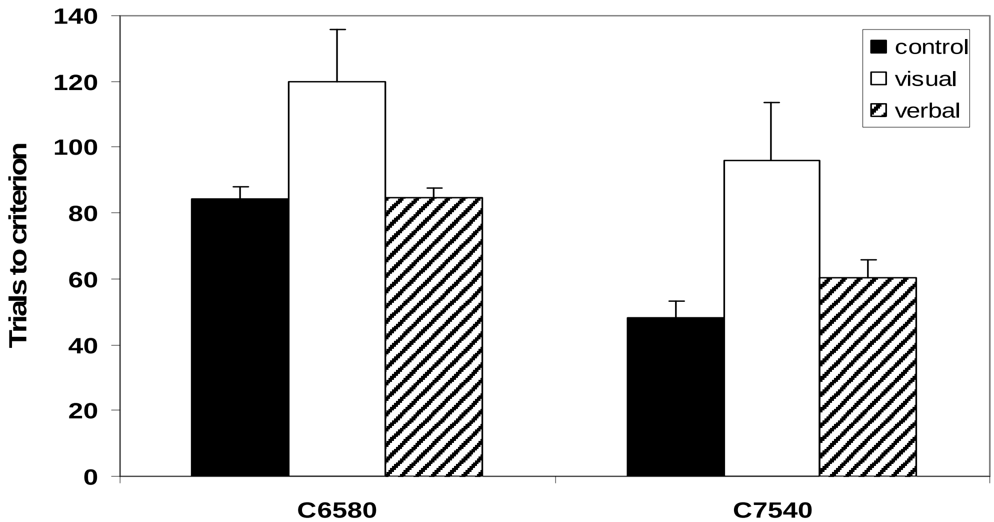

FIGURE 3.8. TRIALS TO CRITERION IN EXPERIMENT 2B. Mean number of trials to reach the criterion of .65 correct over last 80 trials (C6580; left half) and the criterion of .75 correct over last 40 trials (C7540; right half). Error bars denote 68% confidence interval (equivalent to a standard error of mean).

Brief summary

To summarize, we examined visuo-spatial and verbal working memory effects in rule-based category learning when the criterion to be learned was on a cardinal orientation. The results differed from those obtained for a formally identical task that used an oblique orientation as the criterion. We found that the task was much easier to learn and that neither the visuo-spatial nor the verbal working memory task had significant effect on the proportion of learners or final (asymptotic) accuracy. We found adverse effects of both the visuo-spatial and the verbal working memory tasks on mean accuracy during the early stages of learning but we found a significant effect of the visuo-

[13] For an interested reader: The general pattern of results remains similar even if we exclude outliers (so the effect is not driven by extreme values) or include non-learners by assigning them 320 trials to criterion (maximum possible).

spatial, but not the verbal working memory task on the speed of learning as measured by the mean number of trials to reach an accuracy criterion.

How do these results address the hypotheses outlined in the introduction? First, the control data support the hypothesis that cardinal orientations are special both perceptually (leading to higher asymptotic accuracy for learners) and conceptually (leading to high proportion of learners, even early in the experiment). Second, the difference between the visuo-spatial and verbal working memory effects suggests that they may affect different processes associated with rule-based learning. The minimal effect of the verbal working memory task fits with the hypothesis that cardinal orientations are conceptually special, as supported by the control data. The high salience of the cardinal orientation rule leads participants to select the rule early in learning, bypassing much of the working memory demanding hypothesis testing process. The effect of the visuo-spatial working memory task was larger and significantly affected the speed of learning. These results are not likely if the visuo-spatial working memory task was primarily affecting domain non-specific resources (central executive), because the results would be similar to those from the verbal condition. In addition, these results are not likely if visuo-spatial working memory is only involved in learning and representation of the optimal categorization criterion. Because cardinal orientation seems conceptually special (from the control and the verbal condition data), cardinal orientation criterion does not need to be learned in the same way as an oblique orientation and thus should require minimal visuo-spatial working memory resources. Rather, the results seem in accordance with the hypothesis that the visuo-spatial working memory task disrupts the analytic perception of stimuli and leads the participant to take longer to notice the variation of the stimuli around the cardinal orientation. However, as the first attempt to address the visuo-spatial working memory role in category learning, this notion needs to be taken as a working hypothesis only. We elaborate on this possibility below.

General Discussion

Many categorization theories assume that an effortful, working memory demanding process of hypothesis testing is involved in at least some types of category learning (e.g. Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Bruner, Goodnow, & Austin, 1956; Erickson & Kruschke, 1998; Feldman, 2000; Nosofsky, Palmeri, & McKinley, 1994). Previous research (Maddox, Ashby, Ing, & Pickering, 2004) and our Experiment 1 showed that a sequentially presented verbal or visuo-spatial working memory task disrupt rule-based learning, confirming the existence of the effortful hypothesis testing based category learning. Importantly, Maddox et al. (2004) and Experiment 1 also demonstrated that effortful, working memory demanding hypothesis testing is not the only existing process of category learning, because information-integration category learning was not affected at all by a secondary verbal (Maddox, Ashby, Ing, & Pickering, 2004) or visuo-spatial task (present Experiment 1). The lack of the visuo-spatial working memory task effect on information-integration category learning is nontrivial, as both tasks use visual stimuli and are thought to rely on basal ganglia (Filoteo, Maddox, Salmon, & Song, 2005; Lawrence, Sahakian, Hodges, & Rosser, 1996; Lawrence, Watkins, Sahakian, Hodges, & Robbins, 2000; Maddox & Filoteo, 2001; Packard & McGaugh, 1992; Poldrack, 2002; Postle, Jonides, Smith, & Corkin, 1997). Based for example on the COVIS theory of category learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998), we speculate that visuo-spatial working memory and information-integration learning may rely on different subregions of the caudate: visuo-spatial working memory on the head of the caudate nucleus (R. Levy, Friedman, Davachi, & Goldman-Rakic, 1997) while information-integration category learning on the posterior caudate nucleus (Nomura et al., 2007; Seger & Cincotta, 2005). The data from Experiment 1 indeed indicate that there is no interference between the two (at least with respect to the current tasks). Taken together, these results support the notion that both visuo-spatial and verbal working memory tasks impact the processes involved in the hypothesis testing system that mediates rule-based category learning, but neither visuo-spatial nor verbal working memory is crucial for procedural system that mediates information-integration category learning.

Experiment 2 tested the generality of working memory effects in rule-based learning by investigating the effects of a visuo-spatial and a verbal working memory task on rule-based category learning when the criterion was on an oblique orientation (70 degrees) or on the cardinal orientation (90 degrees) of a Gabor stimulus, instead of on spatial frequency. Experiment 2 demonstrated that not all rule-based categories are treated equally. When the criterion was on an oblique orientation, the results replicated those from Maddox et al. and Experiment 1 for which the criterion was on spatial frequency. When the criterion was on a cardinal orientation, we found faster learning with a higher asymptotic accuracy in the control condition. This result confirmed that cardinal orientations are perceptually special (as has been previously established, see e.g.Campbell & Kulikowski, 1966; Furmanski & Engel, 2000). Furthermore, this result suggested that cardinal orientations are also conceptually special, meaning that cardinal orientations constitute salient, spontaneously used categorization boundaries that create intuitive concepts (e.g. right-tilted vs. left-tilted in the present study). Additionally, when the criterion was on a cardinal orientation, we found significantly slower learning when the visuo-spatial task was present, but minimally slower learning when the verbal working memory task was present. These results suggest that visuo-spatial and verbal working memory effects on rule-based learning may be due to dissociable mechanisms. In the remainder of the General Discussion, we will discuss these issues in more detail.

Category Structure Effects on the Distribution of Accuracy Scores

One important result observed in the present studies was the existence of qualitatively different distributions of accuracy scores for the rule-based and the information-integration category structures (see e.g. Figure 3.2). Whereas the accuracy scores in the information-integration conditions were normally distributed, a bimodal distribution was observed in the rule-based conditions, with one modus at chance and a second modus at a high level of accuracy. This pattern held regardless of whether a secondary working memory demanding task was present or absent, and replicates the pattern observed in Chapter 2 that used a dual-task procedure.

These findings provide evidence that rule-based and information-integration category learning is mediated by separate systems. We argue that rule-based category learning is mediated by a hypothesis-testing system whose processing is effortful and attention demanding. Hypothesis-testing systems of this sort are known to have an all-or-none characteristic to their learning that has been studied since the 1960s (Bower & Trabasso, 1963; Trabasso & Bower, 1964). The bimodality observed in our rule-based score distributions is consistent with this hypothesis. We also argue that information-integration category learning is not mediated by a hypothesis-testing system, and instead is mediated by a procedural-based system whose processing is automatic and does not require attention. The fact that the information-integration score distributions were normally distributed and were not affected by a secondary working memory task follows from this hypothesis.

Working Memory Task Effects on Rule-based Accuracy Score Distributions

As outlined earlier, the results from Experiment 1, along with those from Maddox et al (2004) suggest that a sequential verbal or visuo-spatial working memory task has no effect on the distribution of information-integration scores, but has a large effect on the distribution of rule-based scores. The same effect on rule-based learning holds in Experiment 2A that focused on a rule with an oblique orientation criterion. In this section, we elaborate on the nature of the working memory effect on rule-based learning, leaving a discussion of the “cardinal orientation” results from Experiment 2B for later.

As outlined in the previous section, the distributions of accuracy scores in the rule-based conditions were bimodal, with one modus at chance (nonlearners, .50 correct) and a second modus at a high level of accuracy (learners, .80 correct in Experiment 1 and .90 correct in Experiment 2A). One of the most interesting findings from the present study was the fact that the accuracy achieved by learners remained constant across the control and working memory conditions. Rather, the effect of the working memory task was to increase the proportion of participants who fell under the chance (nonlearner) modus relative to the proportion of participants who fell under the higher accuracy (learner) modus. Thus, both the visuo-spatial and verbal working memory tasks seemed to disrupt primarily the process of rule discovery leading fewer participants to discover the correct rule. However, once the correct rule was discovered, participants were just as accurate in applying the rule as learners in the control condition. This finding is important, especially given the general focus on learning curves in category learning research. The typical interpretation of a performance deficit is to assume a *shallowing* of the learning curve and thus a general effect on *average* performance. What the current data suggest is that the effect is not general, but rather increases the probability that a participant will fail to discover the correct rule.

Interaction between the hypothesis testing and procedural system under secondary task

One reasonable prediction from the COVIS model (Ashby, Alfonso-Reese, Turken, & Waldron, 1998), and perhaps also implicit expectation in other multiple system models of learning, is that the procedural system would take over and dominate rule-based category learning when the hypothesis testing system is disrupted by a working memory task. COVIS assumes that both category learning systems (the hypothesis testing system and the procedural system) aim to learn every categorization task and that the two systems compete to generate the response on each trial. If the procedural system is unaffected by the presence of a working memory task then it should dominate the hypothesis testing system. This did not seem to be the case in the current studies because we found no compensation in the rule-based task by the procedural system – the accuracy scores remained low and were not normally distributed – when the working memory task was present. In Chapter 2 (Zeithamova & Maddox, 2006), we similarly found that participants learning a more complex conjunctive rule-based categorization task under the dual Stroop task interference tended to use (rule-based) unidimensional strategies or guessing rather than converting to information-integration strategies.

There are at least two possible explanations for this finding. First, the current rule-based category structures may be unfavorable for learning in the procedural system; for instance, because of their relatively high intra-category variability and low inter-category variability. Second, the procedural system may be learning the task, but the participant may be highly biased towards the (unsuccessful) hypothesis testing system by the secondary working memory task. COVIS assumes that there is an initial bias toward the hypothesis-testing system. Participants may never abandon this bias under the secondary working memory task conditions, perhaps because also because processing of the feedback regarding each system's performance is compromised. Given the lack of research that directly examines system level interactions these hypotheses should be considered speculative at this time.

Dissociating Visuo-spatial and Verbal Working Memory Effects on Rule-Based Category Learning

Visuo-spatial and verbal working memory are behaviorally and neurally dissociable (e.g. Baddeley, 1995; Goldman-Rakic, 1998; Shah & Miyake, 1996). It is thus reasonable to consider the possibility that although both visuo-spatial and verbal working memory tasks adversely affect rule-based learning, the locus of their effect may differ. The effect of the verbal working memory task on rule-based category learning reported in Maddox et al. (2004) and replicated in Experiment 2A would be expected by any hypothesis testing model. The effect of the visuo-spatial working memory task on rule-based category learning observed in the present studies is less straightforward, because no existing category learning theory addresses the possible role of visuo-spatial working memory. We speculated that the observed visuo-spatial task effect in rule-based category learning may act indirectly via some kind of general attention or control mechanism common to both visuo-spatial and verbal working memory (e.g. central executive in Baddeley & Logie, 1999) or may be mediated by a different, independent mechanism. Two pieces of evidence argue against the notion of an indirect effect via commonly shared resources and instead support the notion of two different mechanisms for visuo-spatial and verbal working memory task effects. We will now discuss the two pieces of evidence in more detail.

Comparison of Maddox et al. (2004) and Experiment 1

Although qualitatively similar, the verbal working memory task effect reported in Maddox et al. (2004) seem to differ in magnitude from the visuo-spatial working memory task effect found in Experiment 1. To compare the working memory effects on rule-based learning across the two experiments, we computed effect sizes (Cohen, 1988)[14] for the final block performance drop in each experimental condition compared to the associated control condition (Figure 3.9, left panel). These analyses should be interpreted with caution bearing in mind that Cohen's effect size measure is derived from means and the means in these experiments represent a relative mixture of two populations of participants rather than an average participant. Nevertheless, these analyses are suggestive and seem to shed some light on the nature of the working memory effects. As is apparent from Figure 3.9, left panel, the verbal working memory task had a large effect when immediately following categorization feedback, but only a small effect when participants were first allowed to process the categorization feedback for 2500 ms. The visuo-spatial working memory task had an intermediate effect when immediately following categorization feedback, but continued to disrupt performance even when the participant was first allowed to process the categorization feedback for 2500 ms. Although by itself this finding is inconclusive, it favors the notion that verbal and visuo-spatial working memory tasks have at least partially dissociable effects on rule-based category learning.

[14] Effect size measures the magnitude of an effect independent of sample size (number of participants in a study). A frequently used measure of effect size is Cohen's D that is computed as the difference in means divided by the pooled standard deviation (d prime). As a rule of thumb, D above .2 is considered a small effect, D above .5 is considered a medium effect and D above .8 is considered a large effect (Cohen, 1988).

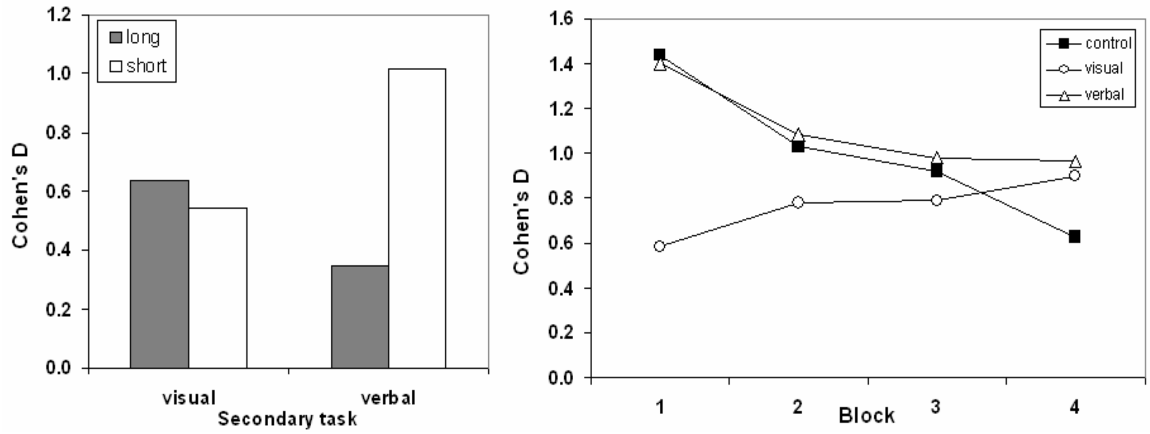

FIGURE 3.9. EFFECT SIZE COMPARISON OF THE SECONDARY VISUO-SPATIAL AND VERBAL WORKING MEMORY TASK. Left panel: Effect size (rule-based category learning performance decrement in Cohen's d) of a secondary visuo-spatial or verbal working memory task following categorization task feedback after 2500 ms delay ("long:" black bars) or immediately ("short:" gray bars). Data for visuo-spatial working memory effects are from Experiment 1, data for verbal working memory effects were computed from Maddox et al. (2004). Right panel: Effect size of moving criterion from 70 to 90 degrees. Data were obtained by comparing Experiment 2A and Experiment 2B.

The effect of the cardinality of a categorization criterion

One goal of this Chapter was to investigate the effect of cardinality of a criterion on rule-based category learning. We asked whether cardinal orientations are conceptually special and whether working memory effects observed with a general (arbitrary) criterion rule-based learning replicate for a cardinal criterion. The pattern of results for the cardinal criterion exhibited several differences from those observed when the criterion was on an oblique orientation or on spatial frequency. First, in the control condition, performance reached asymptote much earlier and at a higher proportion correct for learners than for the formally identical structure using criterion on oblique orientation. The results suggest that a criterion on cardinal orientation is easier to learn than a criterion on an oblique orientation not only perceptually, but also conceptually; it seems that cardinal orientation plays a role in cognition as an intuitive, salient categorization rule. Second, the speed of learning was adversely affected only by the visuo-spatial working memory task, with minimal effect of the verbal working memory task. This finding implies that the salience of the cardinal criterion may be occluded by a secondary visuo-spatial working memory task, perhaps because it disrupts memory traces of previously seen stimuli. In Figure 3.9 (right panel), we accentuate this by examining the effect of the cardinality of a criterion throughout the learning process. For each block and each condition, we computed the effect size (Cohen's D) of changing the categorization criterion from 70 to 90 degrees (i.e. the difference between scores in Experiment 2A and Experiment 2B). The advantage for the cardinal orientation was large, mainly early in the experiment and of about the same magnitude for the control and the verbal condition. On the other hand, the advantage for the cardinal orientation was much smaller for the visual condition, especially early in the learning, consistent with the trials-to-criterion analysis. This finding further contributes to the notion that visuo-spatial and verbal working memory effects in rule-based category learning may be due to different mechanisms. Next we discuss the possible role of verbal and visuo-spatial working memory in category learning separately.

The role of verbal working memory in category learning

Previous literature implied verbal working memory in hypothesis generation, selection and testing (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Dougherty & Hunter, 2003). The presented results are all consistent with this notion. An effect of verbal working memory task was observed when participants were presented with two generally equally salient dimensions and set of criterions along those dimensions, but not when a highly salient rule was offered in Experiment 2B. The minimal effect of the verbal working memory task in cardinal orientation rule learning suggests that cardinal orientations are special not only perceptually, but also conceptually (i.e. they serve as highly salient categorization rules). Specifically, the values along the orientation dimension seems to naturally fall into two distinct classes (right tilted stimuli, left tilted stimuli) when orientation varies around a cardinal direction. Thus, verbal working memory may not be needed when selecting and applying a highly salient rule, because much of the effortful hypothesis testing process is skipped.

The role of visuo-spatial working memory in category learning

What may be the possible role of visuo-spatial working memory in rule-based category learning? One notion outlined in the introduction was that visuo-spatial working memory may be needed to represent the optimal *categorization criterion*, while the verbal working memory may be needed to represent the *categorization rule*. In other words, visuo-spatial working memory may be important in representing the criterion of 70 degrees (in Experiment 2A) or 90 degrees (in Experiment 2B), but may not be needed to represent the rule: "respond A if the orientation is greater than the criterion and respond B if the orientation is less than the criterion". The proportion of learners is a measure of the proportion of participants that discover the correct *rule*. The specific accuracy of these learners depends on how well they learn and remember the optimal *criterion*, i.e. find a particular value of the orientation that best discriminates between the high orientation and the low orientations. If visuo-spatial working memory is needed only to hold the optimal categorization criterion, but is not crucial in the process of rule discovery, we would expect the proportion of learners in Experiment 1 and Experiment 2A to be about the same in the visual condition as in the control condition, but for their accuracy to be lower, due to noisier categorization criterion representation. Contrary to this hypothesis, we observed a lower proportion of learners, but their accuracy to be about the same as in the control condition. In Experiment 2B, we would expect minimal effect of the visuo-spatial working memory task, because representing a criterion on cardinal orientation – a highly learned natural boundary – should require minimal working memory resources. Contrary to this prediction, we found an adverse effect of the visuo-spatial task on the speed of learning.

Another notion was that visuo-spatial working memory may be needed for analytic evaluation of a stimulus and its individual dimensions. The present results are quite consistent with this notion. If the visuo-spatial working memory task disrupts analytical perception of the stimuli, it may make it difficult to identify the dimensional structure of the stimulus and thus to generate possible rules in Experiment 1 and Experiment 2A. Disrupting analytic processing of the stimuli may also disrupt the perception of variation around a cardinal orientation, perhaps by disrupting the visual memory representation of previous stimuli, and thus delaying the time when the rule on cardinal orientation may be selected. On the other hand, the secondary visuo-spatial working memory task had no effect on information-integration category learning in Experiment 1 because information-integration category learning requires holistic rather than analytic perception of a stimulus. Although consistent with the presented results, the proposed role of visuo-spatial working memory in analytic evaluation of categorization stimuli remains speculative until further research addresses this issue.

Summary

The results presented in this Chapter extend our understanding of the role of working memory in category learning by examining the effects of sequentially presented visuo-spatial and verbal working memory tasks on rule-based and information-integration category learning. In line with the results from Maddox et al. (2004) that used a verbal working memory task, we found no effect of a visuo-spatial working memory task on information-integration category learning, but a significant effect on rule-based category learning when the categorization criterion was on spatial frequency of a Gabor stimulus. We also replicated the effect of both visuo-spatial and verbal working memory tasks on rule-based learning with a categorization criterion on an oblique orientation of a Gabor stimulus. These results add to the evidence for the existence of multiple category learning mechanisms. When examining the effect of the secondary tasks on rule-based learning in more detail, we interestingly found that the presence of both visuo-spatial and verbal working memory tasks affected primarily the proportion of participants who discovered the rule by the end of training; the accuracy of those learners remained the same across conditions. A different pattern of working memory effects was observed when the rule-based categorization criterion was on a cardinal orientation. We found a minimal effect of the verbal working memory task, but a large effect of the visuo-spatial working memory task on the speed of learning. These results suggest that the cardinal orientation serves as a highly salient, natural categorization boundary and that visuo-spatial and verbal working memory effects on rule-based category learning are at least partially dissociable. A plausible role for visuo-spatial working memory consistent with the presented results is analytic evaluation of individual stimulus dimensions.

SUPPLEMENTAL DATA. DESCRIPTION OF THE MODEL FITTING PROCEDURE APPLIED TO THE SCORE DISTRIBUTIONS FROM EXPERIMENTS 1 AND 2A

A visual inspection of the accuracy score distributions from the Experiment 1 and 2A rule-based groups suggested that these distributions are bimodal (see Figure 3.2, left panel and Figure 3.5, left panel). In this section, we describe a method for characterizing the nature of these bimodal distributions and the effect that the verbal and visuo-spatial working memory tasks had on these distributions.

We fit a series of five models of various degrees of generality to the distribution of accuracy scores from the final block of trials, separately for each condition. After best fitting parameters for each model were estimated using maximum likelihood method, we used the BIC measure (Schwarz, 1978) to compare the models and to determine the best fitting model.

Model description

Model 1: Model 1 is the most general model. This model assumes that the score distribution is bimodal, with each modus being best described by a normal distribution. This model has five free parameters: the mean and standard deviation of the first normal distribution, the mean and standard deviation of the second distribution, and the relative weight of the second distribution (with the relative weight of the first distribution being 1-relative weight of the second distribution). The relative weight of each distribution represents the proportion of participants whose accuracy scores contributed to that distribution (modus).

Model 2: Model 2 is a special case of Model 1 for which the mean of one distribution is fixed at chance (.5). This model instantiates the hypothesis that one group of participants did not discover the correct rule and that their mean accuracy is at chance level of .5. This model has four free parameters.

Model 3: Model 3 is a special case of Model 2 for which the chance distribution is assumed to be binomial (rather than normally distributed), with the probability of success being .5 on each of the 80 trials that constituted the final block. Because the standard deviation of a binomial distribution is derived from the mean and the number of trials, there are no free parameters associated with the "chance distribution" leaving the three free parameters associated with the second distribution to be estimated.

Model 4: Model 4 is a special case of Model 3 for which the non-chance distribution is also assumed to be binomial. This model has two free parameters: the mean and relative weight of the second distribution.

Model 5: Model 5 was included as an additional check of the Kolmogorov-Smirnov test of normality. This model assumes that the distributions of scores are best characterized by one normal distribution. This model has two free parameters: the mean and standard deviation of the normal distribution. This model provided poor fits in all cases.

Results

Best fitting models in Experiment 1.

RB control – Model 3:

Distribution 1: binomial (mean = .5, relative weight = .318)

Distribution 2: normal (mean = .814, sd = .065, relative weight=.692)

RB long – Model 4:

Distribution 1: binomial (mean = .5, relative weight = .560)

Distribution 2: binomial (mean = .788, relative weight = .440)

RB short – Model 4:

Distribution 1: binomial (mean = .5, relative weight = .512)

Distribution 2: binomial (mean = .786, relative weight = .488)

Best fitting models in Experiment 2A.

Control – Model 3:

Distribution 1: binomial (mean = .5, relative weight = .167)

Distribution 2: normal (mean = .897, sd = .046, relative weight=.833)

Visual – Model 3:

Distribution 1: binomial (mean = .5, relative weight = .500)

Distribution 2: normal (mean = .907, sd = .058, relative weight=.500)

Verbal – Model 2:

Distribution 1: normal (mean = .5, sd = .032, relative weight = .500)

Distribution 2: normal (mean = .879, sd = .058, relative weight=.500)

Chapter 4: Prototype learning is not a uniform process

In the previous two chapters, we focused on one family of category learning tasks, demonstrating dissociation between the rule-based task and the information-integration task. We concluded that the rule-based task is supported by frontally mediated reasoning system while the information-integration task is supported by striatum mediated procedural learning system. We will now transition into a domain of another type of categorization task – the prototype learning task – that has been traditionally regarded as a third type of task, relying on yet different, perceptual learning system. Chapter 4 will review the basic concepts from the prototype literature and propose dissociation within a realm of prototype learning.

Recall that prototype is a collection of characteristic features of a category, or the ideal exemplar of the category. Prototypes are thought to provide the abstract representation for many natural categories and concepts (Rosch, 1973, 1975b; Rosch & Mervis, 1975). Prototype theory assumes that the central tendency (or prototype) of a category is abstracted during encounters with category exemplars. Category members are centered around prototypes based on family resemblance principle (Wittgenstein, 1953), meaning that most of them share a number of characteristic features (like members of a family), but none of the features is necessary or sufficient for category membership. While category membership in formal concepts is all-or-none, based on presence or absence of defining features, category membership in natural concepts have been shown to be graded, based on the comparison of an instance to the category prototype (McCloskey & Glucksberg, 1978; Rosch, 1973, 1975b). The closer an instance matches the category prototype, the faster and more reliable it can be verified as a category member. For example, people are faster to verify that "Robin is a bird" than "Penguin is a bird" (Rosch & Mervis, 1975) and agree more reliably with others and with themselves over time that “apple is a fruit” than “pumpkin is a fruit” (McCloskey & Glucksberg, 1978).

Prototype effects have been repeatedly demonstrated for novel concept learning as well. In their seminal work, Posner and Keele (1968; , 1970) used dot patterns as prototypes and their distortions as category exemplars. A number of studies followed them and showed that participants were able to classify unseen prototypes with higher accuracy than other category exemplars, and the prototypes were less susceptible to forgetting after a one week delay than the actual training patterns (Homa & Little, 1985; Posner & Keele, 1970; W. Strange, Keeney, Kessel, & Jenkins, 1970). Reed (1972) was one of the pioneers who explicitly compared models that assume prototype abstraction with a number of other models proposed for classification learning and found that the prototype model outperforms other models when subjects are presented with the category exemplars sequentially, limiting thus individual exemplars availability during retrieval. Reed suggested that prototype representation optimizes cognitive economy when memory limitations come to play.

Alternative accounts of prototype effects have been proposed as well. The most prominent is exemplar theory (Medin & Schaffer, 1978; Nosofsky, 1986, 1988; Nosofsky, Clark, & Shin, 1989; Nosofsky, Kruschke, & McKinley, 1992; Zaki & Nosofsky, 2004). Exemplar theory assumes that in the process of category learning, people represent and store each encountered category exemplar. In other words, categories are represented by all their exemplars rather than a single prototype. The prototype – the central tendency of the category – not only does not serve as a representation of the category, it is actually never abstracted. In the process of categorization, a novel stimulus is compared to the retrieved exemplars from all relevant categories and assigned to a category based on the summed similarity of that stimulus to the category exemplars. The reason why prototypes are categorized faster and more accurately than non-central exemplars is because they are more similar to the stored exemplars of a category. Although exemplar theory does not address the question of the neural substrate supporting the exemplar representation, the medial temporal lobe would be the likely candidate to support exemplar memorization.

Prototype and exemplar models do not stand sharply against each other. Both prototype and exemplar models can account well for a wide range of empirical data. In general, exemplar models provide better fits when training involves smaller number of stimuli, small number of stimulus dimension, and frequent exemplar repetitions; prototype models provide better fits when training involves larger number of stimuli, that vary along several dimensions and are presented infrequently (Minda & Smith, 2001; J. D. Smith & Minda, 1998). Newer "clustering" models propose that category knowledge is represented at an intermediate level of abstraction, with exemplar clusters being more abstract than individual exemplars, but less general than a single central prototype (Love, Medin, & Gureckis, 2004; Verbeemen, Vanpaemel, Pattyn, Storms, & Verguts, 2007). Moreover, if exemplar representations are stored across distributed neural network, the clusters and prototypes arise from the network as an emergent property, and the exemplar, clustering and prototype models become equivalent to each other, differing only in the level of description. For the purpose of this dissertation, we will assume that prototype abstraction is a cognitive process involved in novel concept learning as well as a part of natural concept representation, leaving out the discussion of its ontological status. In the reminder of this Chapter, we review existing prototype learning studies, propose dissociation between two types of prototype tasks, and discuss limits of the existing literature in addressing this dissociation. Chapters 5 and 6 will then explore cognitive and neural properties of two prototype learning tasks using a novel methodology overcoming these limits.

VERSIONS OF THE PROTOTYPE LEARNING STUDIES

When exploring prototype literature, we realize that prototype learning studies come in various flavors. One of the most pronounced differences across studies is between a multiple category/multiple prototype learning task, here referred to as an "A/B task", and a one category/one prototype learning task, here referred to as an "A/nonA task”. In the A/B task, participants learn to classify exemplars into two (or more) contrasting categories. This version may be representative of concept learning in children when a parent may walk around with a child, pointing out different exemplars from different concepts. “Look, this is a cow, this is a horse, this is a goat.” In the A/nonA task, only one category exists and participants learn to classify exemplars as members or nonmembers of that category. This version may be representative of concept learning in children based on an exposure to a large number of exemplars of one concept. “Look, a flock of chickens. These all are chickens.”

In most studies, the information about which prototype task (A/B or A/nonA) was used is typically buried deep in the method section and conclusions derived from one version of the prototype task are readily generalized to the other version. However, one could argue that structure of these two task types may yield recruitment of different cognitive and neural processes. In the A/nonA task, participants are likely to form a representation of a single prototype and then compare each test item to this single prototype. If the new stimulus is sufficiently similar to the prototype representation, it will be endorsed to the category; otherwise it will be categorized as a non-member. Novelty or familiarity signals from early processing areas may be used as a basis for successful categorization. In the A/B task, participants are likely to form representations of two distinct categories centered on two prototypes. Each new stimulus is then compared to both of these prototypes and endorsed to the category of the prototype which is closer to the current stimulus. Familiarity or novelty signals are not sufficient for successful performance.

Indeed, some evidence suggests that the processes involved in these two types of tasks differ. First, a few behavioral studies compared A/B and A/nonA task directly. Goldstone and colleagues (Corneille, Goldstone, Queller, & Potter, 2006; Goldstone, 1996; Goldstone, Steyvers, & Rogosky, 2003) compared prototype representations that participants acquired in the A/B task with those acquired in the A/nonA task. In the A/B task, symmetric representations of two contrasting categories were formed and features that best differentiated between categories were emphasized. Participants found caricatures – stimuli that overemphasized the distinctive feature – as the most typical exemplars of the categories. In the A/nonA task, representation of the categories was asymmetric and category A prototype included all characteristic features of category A, whether or not they were diagnostic for category membership. Prototypes, rather than caricatures, were found to be the most typical exemplars of the category. Another study that directly compared behavioral performance in the A/nonA task with the A/B task was by Casale and Ashby (in press). They compared performance in the A/B version and the A/nonA version of the traditional dot pattern task. In Experiment 1, they found that learning in the A/nonA task is more sensitive to the exemplar distortion level than is in the A/B task, but they failed to equate baseline accuracy in the two tasks, so the conclusion should be taken with caution. In Experiment 2, they found that A/B performance is highly affected by a removal of feedback while A/nonA performance is less affected. They argued that the A/nonA task is supported by the perceptual representation system (Schacter, 1990), a learning system that plays a role in perceptual priming (Wiggs & Martin, 1998), while the A/B task is supported by a different (unspecified) learning system.

Second, indirect evidence comes from the neuropsychological literature. Prototype learning task has been a traditional example of nondeclarative learning, intact in amnesic patients (Knowlton & Squire, 1993). This view has been recently challenged by Zaki and colleagues (Zaki, Nosofsky, Jessup, & Unversagt, 2003) that argued that prototype task appears intact because it is too simple (ceiling effect), but becomes impaired when made more challenging. As the simple version, Zaki and colleagues used the A/nonA task, as was used in the seminal Knowlton and Squire (1993) paper; for the challenging version, they asked participants to learn two categories, or what we call A/B task. Indeed, the simple (A/nonA) task was not impaired while the challenging task (A/B) was. Alternatively, this dissociation may be based on the qualitative rather than quantitative differences between the two tasks. Ashby and colleagues (Ashby & Maddox, 2005; Ashby & O'Brien, 2005) noted that impaired prototype learning was reported in amnesia when the A/B task was used (Zaki, Nosofsky, Jessup, & Unversagt, 2003) while intact prototype learning was reported in amnesia when the A/nonA task was used (Bozoki, Grossman, & Smith, 2006; Keri, Kalman, Kelemen, Benedek, & Janka, 2001; Knowlton & Squire, 1993). They suggested that A/B learning is mediated by declarative memory processes while A/nonA learning is mediated by nondeclarative, perceptual learning processes.

Third, a wide range of results reported in the neuroimaging literature also indicates that prototype learning is not a single process. Some studies revealed dependence on brain networks associated with episodic memory (Reber, Gitelman, Parrish, & Mesulam, 2003), while others revealed involvement of other networks, including those associated with perceptual learning (Reber, Stark, & Squire, 1998a), visuospatial attention (Little & Thulborn, 2005), and visual reasoning (Seger et al., 2000). However, unlike the results from learning in amnesia, the results from neuroimaging studies do not become cohesive even when task type is taken into account. For example, medial temporal lobe involvement has been reported both in some studies that employed A/B task (DeGutis & D'Esposito, 2007; Little, Shin, Sisco, & Thulborn, 2006), and in some studies that employed A/nonA task (Aizenstein et al., 2000; Reber, Gitelman, Parrish, & Mesulam, 2003). Thus, the nature of the dissociation between A/B and A/nonA prototype learning remains unanswered.

LIMITED COMPARABILITY ACROSS EXISTING STUDIES

Direct comparison of the results from the A/B and A/nonA experiments in the existing literature is complicated by a number of frequent differences in both methodology and data analysis applied. First, learning mode often differs between the two tasks. The A/B task always involves intentional learning – the participants are aware of the goal of the experiment from the beginning and intentionally try to learn the characteristics of the categories based on presented category labels and/or corrective feedback. In the A/nonA task, learning is often incidental – participants passively view category exemplars during the study phase without being aware of the goal of the experiment. After that, participants are instructed that all the previously viewed exemplars come from a single category and are asked to classify new items as either coming from the same category or not.

Intentionality of learning has a profound effect on what kind of categories are typically learned. In her influential work, Kemler-Nelson (1984) found that people typically represent learned categories in terms of criterial attributes or rules when learning intentionally while acquire family-resemblance-based representation when learning is incidental. Two fMRI studies compared neural activation in both incidental and intentional version of an A/nonA task (Aizenstein et al., 2000; Reber, Gitelman, Parrish, & Mesulam, 2003). Aizenstein and colleagues found decreased activation in occipital cortex for prototypical ("A") versus non-prototypical ("non-A") items in the incidental version while increased activation in the occipital cortex, and decreased activation in medial temporal lobe, parietal lobe and frontal regions in the same contrast during the intentional version of the task. Reber and colleagues also found occipital deactivation in the incidental task, and increased activation in prefrontal cortex, occipital cortex and precuneus during intentional version. Given the importance of the learning mode, it is not obvious whether this factor alone cannot account for the differences observed between the studies that employ the (intentional) A/B task versus the (often incidental) A/nonA task.

Second, category structures typically used differ between the two tasks. In the A/B task, each category is generated as a collection of distortions from a category prototype, with prototypes varying between categories. Each category has an internally consistent structure, forming a coherent cluster centered in the perceptual space around the category prototype. In the A/nonA task, as the name suggests, only the A category items are generated by a distortion from a prototype, forming a coherent cluster. The non-A items are typically randomly generated stimuli that have no consistent structure and are only defined negatively, as items that are not distortions from the A prototype. This difference between A and non-A stimuli constitutes a potential flaw of the A/nonA studies that argue for implicit learning in amnesia. Category A exemplars are necessarily perceptually more similar to each other than are the non-A items to each other. Exemplar similarity to recent stimuli have been show to strongly influence category performance (A. S. Levy & Heshka, 1973) and similarity among category A exemplar against the background of unrelated non-A stimuli may support unintended "learning" during transfer phase (Zaki & Nosofsky, 2007). If we are to make the distinction between the A/B task and the A/nonA task, we need to eliminate the confounding factor of stimulus differences.

Third, in the neuroimaging literature, different fMRI contrasts have been used in A/B studies and A/nonA studies. In A/B studies, the BOLD signal was typically contrasted between the prototype task and fixation baseline (Little & Thulborn, 2005; Seger et al., 2000), whereas in A/nonA studies, the BOLD signal was contrasted between categorical (A) items and non-categorical (nonA) items, both being a part of the prototype task (Reber, Stark, & Squire, 1998b; Reber, Wong, & Buxton, 2002). Even if both tasks were supported by identical brain areas, performing identical functions, the difference in contrasts used could by itself produce differential patterns of activation. For example, areas that may support performance in the A/nonA task, but do not directly differentiate between A items and nonA items, would be subtracted out in the A versus nonA contrast, but may be identified in the A/B task versus baseline contrast. Additionally, the differential activation to categorical versus noncategorical items in the traditional A/nonA task may reflect simple sequential contrast effect (as categorical items are more likely to be similar to their immediate predecessor), rather than reflecting category membership. Again, the conclusion from the existing studies is limited as long as actual stimuli and fMRI constrast are not equated.

OVERVIEW OF THE STUDIES IN CHAPTER 5 AND CHAPTER 6

The goal of Chapter 5 and Chapter 6 is to explicitly study similarities and differences between the multiple category learning (A/B) task and the one-category learning (A/nonA) task using behavioral and neuroimaging methods. For a more reliable comparison, we eliminate the common external differences between the two tasks outlined above. First, we use intentional learning mode in both tasks. Besides equating the learning mode, this manipulation will allow us to test individual participants repeatedly and use within subject comparison in the functional MRI study presented in Chapter 6. Second, we use identical underlying category structures in both tasks. An A/B task can be converted into an A/nonA task by referring to “B” items as “non-A” items. The meaningfulness and effectiveness of such manipulation was shown by Goldstone (1996). Goldstone tested whether asymmetric category representation, typical for the A/nonA task, can be achieved with A/B category structures when the category B label is simply changed to nonA label. Using a simple change in instructions (e.g. “Press “N” for a painting by Noogan” in the A/B task versus “Press “N” for a painting NOT by Yarpleaux” in the A/nonA task), he found that resulting category representation was similar to that normally found in the A/nonA task. Using identical category structures for a direct comparison of the two tasks is especially important in neuroimaging studies, as stimulus effects may drive any observed differences between the two tasks. No neuroimaging or neuropsychological study however employed the label-only manipulation before. Last but not least, we will use a common baseline and a direct contrast in our fMRI study (Chapter 6) to identify commonalities and differences between neural activation in the two tasks. In all of the studies, we will use novel category structures that are based on family resemblance, to mimic a character commonly attributed to natural categories.

In Chapter 5, we will examine basic behavioral characteristics of four different versions of the prototype learning task used in the literature – two versions of the A/B task and two versions of the A/nonA task – while equating category structures and stimuli used in each. We will demonstrate that various flavors of the prototype learning task can all show typical prototype effects and can be on average about equally difficult. In Chapter 6, we will examine the neural underpinnings of the A/B task and the A/nonA task in an fMRI study, using methodology that holds constant the category structures, learning mode, and the fMRI contrasts. Behavioral data will show that the performance profile in the within subject design replicates the results from the between subject design, demonstrating viability of our ambition to directly compare neural activation in the two tasks within subject. Additionally, we will show that the two tasks are poorly correlated within subject, providing first behavioral evidence that they may be supported by different cognitive processes. Neuroimaging data will show that the A/B task and the A/nonA task involve both common and dissociable neural regions, with the dissociation mapping relatively well, but not perfectly, on the declarative versus nondeclarative distinction suggested by previous neuropsychological studies.

Chapter 5. Exploring prototype learning tasks: Constant category structures in different training variants

In the previous chapter, we discussed two versions of the prototype learning task, the A/B task and the A/nonA task. A proposal has been made that processes involved in the performance of the two tasks may differ, but a number of methodological issues limit the conclusions that we can do based on the existing literature. The most serious one is that prototype effects in the traditional A/nonA task are confounded with exemplar similarity and/or sequential effects. In the A/nonA task, participants first view a set of exemplars that all come from one category (are distortions of a single prototype) and later need to classify new stimuli as members of the category or nonmembers. Traditionally, the non-categorical stimuli are unrelated not only to the original prototype, but also to each other. Because categorical stimuli are similar to each other while noncategorical are not, one may argue that the A/nonA task appears to be spared in amnesia not because any learning has occurred during training, but because participant may rely on working memory during the test and indicate whether or not a stimulus seems similar to one or two previous ones. In order to rigorously compare the A/B task and the A/nonA task, we need to equate category structures used in the two tasks, while preserving characteristics of each task. Chapter 5 has two related goals, to compare behavioral profiles of the A/B task and the A/nonA task when the confounding factors are limited and to develop a new procedure that would allow a direct comparison of the two tasks using functional magnetic resonance (applied latter in Chapter 6). To achieve these goals, we propose a methodology that equates category structures and stimuli between different versions of the prototype task, but preserves the relationship between the categories as they exist in the traditional A/B and A/nonA tasks.

Constant category structure

The first challenge is to find category structures and stimuli that could be meaningfully used in both tasks. In the A/B task, two unique prototypes exists, prototype A and prototype B, and their features are parametrically distorted to create exemplars of category A and exemplars of category B. Prototype theory assumes that a stimulus is assigned to the category whose prototype is closer to the stimuli in the perceptual space (Homa, Sterling, & Trepel, 1981; Posner & Keele, 1970; Reed, 1972). In the A/nonA task, only one prototype exists and its distortions constitute category A exemplars. With dot-pattern and similar stimuli, the non-categorical (non-A) stimuli are generating by random selection of dot locations or by setting some binary features to random values. Prototype theory assumes that a stimulus is assigned to the category A if its distance to the category A prototype is smaller than some threshold criterion (Casale & Ashby, in press; J. D. Smith & Minda, 2002). A graphical depiction of the structure of the two tasks is presented in Figure 5.1. The boundaries between the categories for both A/B and A/nonA tasks are considered "fuzzy" (McCloskey & Glucksberg, 1978, 1979; Roth & Mervis, 1983), meaning that category membership is probabilistic rather than deterministic.

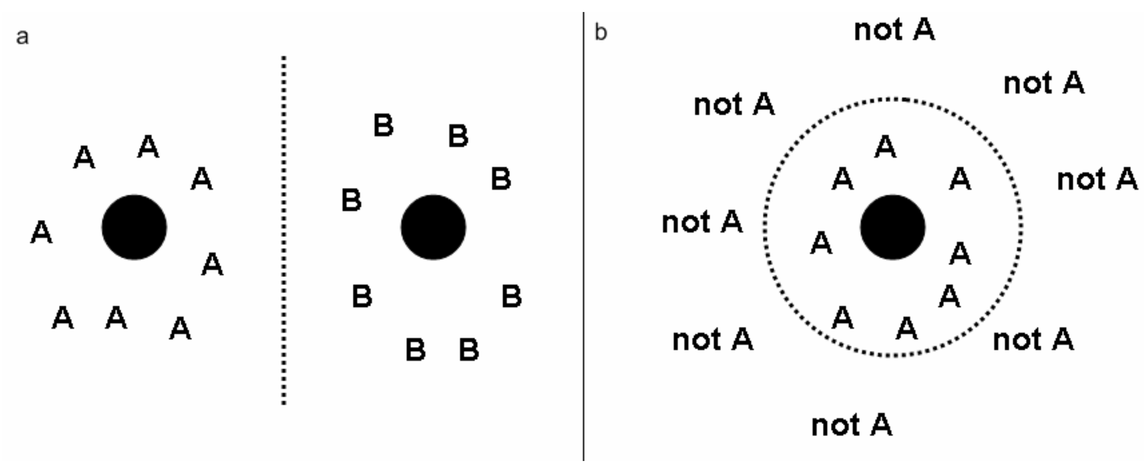

FIGURE 5.1. CATEGORY STRUCTURES USED IN (A) A/B TASKS, (B) A/NONA TASKS. The black circles represent category prototopes, the dotted lines represent "fuzzy" category boundaries.

In order to equate category structures and stimuli in the two tasks, we need to find a structure where category B stimuli are the same as not A stimuli. One possibility is to use traditional A/B category structure, using two unrelated stimuli as category A prototype and category B prototype, but change the label of the category B stimuli into “not A” for the A/nonA task (Goldstone, 1996). However, the specific circular character of the boundary between the A and the non-A space would be lost. One way to preserve the character of the boundary in both tasks is to use a category structure that is like a sphere (Figure 5.2.a). On the sphere, the north and south poles represent the two prototypes, the equator represents the category boundary, and the category membership of an exemplar depends on its latitude on the sphere. The category B and not-category A stimuli are naturally equated as the further one goes away from the North Pole (prototype A), the close one gets to the South Pole (prototype B). Depending on the view, the two categories have a straight or circular boundary (Figure 5.2. b,c). To create such a sphere-like category structure, we use stimuli that vary along multiple binary-valued dimensions and category structures defined based on family resemblance, with one set of values along all dimensions representing category A prototype and the opposite set of values along all dimensions representing category B prototype (Figure 5.3). The binary value dimensions provide a unique opportunity to naturally equate category B members with non-A members, as the farther a stimulus is from the category A prototype, the closer it is to the category B prototype. More details of the stimuli and category structure used are presented in the method section.

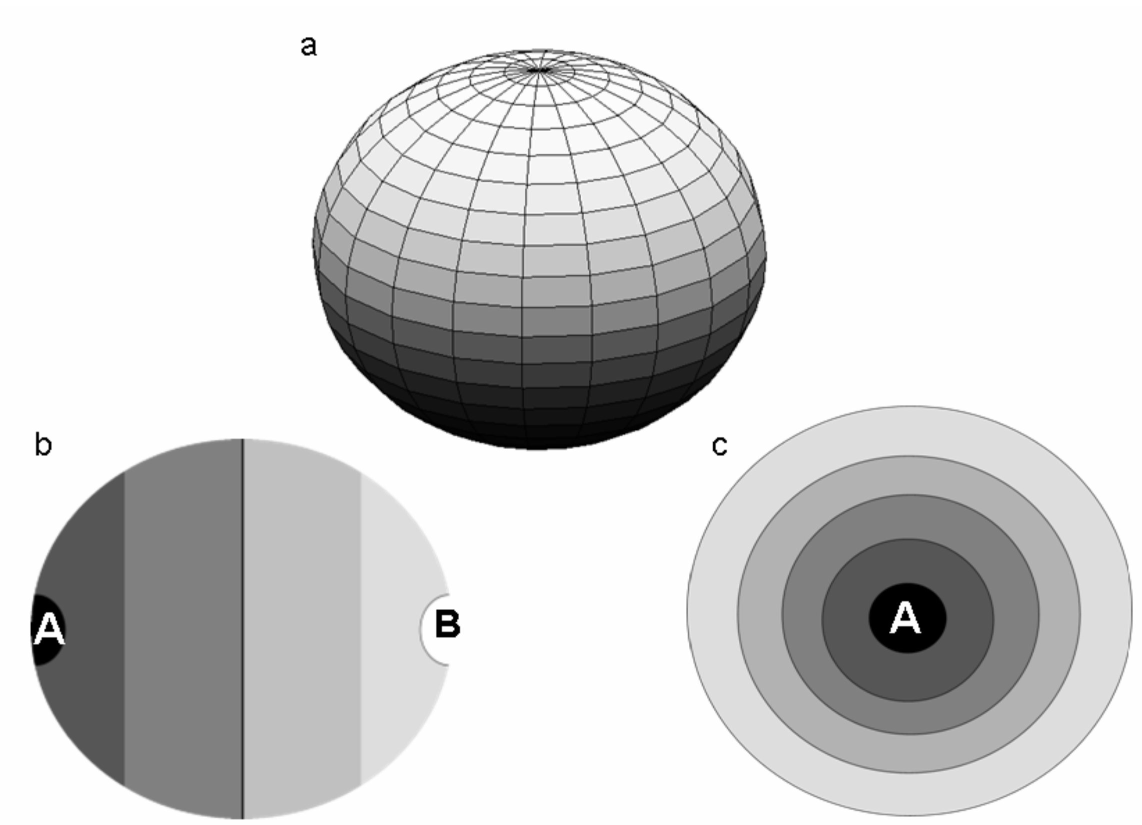

FIGURE 5.2. SCHEMA OF A SPHERE-LIKE CATEGORY STRUCTURE. a) The outside view of the structure; b) The A/B task view of the structure; c) the A/nonA task view of the structure.

Different training variants

Four different versions of a prototype learning task are used in the literature – two variants of the A/B task and two variants of the A/nonA task. The two variants of the A/B task are feedback training A/B task and observational training A/B task. The feedback training version of the A/B task is the most common variant of the A/B task (Little, Shin, Sisco, & Thulborn, 2006; Little & Thulborn, 2006; Minda & Smith, 2001; Posner & Keele, 1968, 1970; Reed, 1972; Seger et al., 2000), requiring the participant on each trail to indicate category membership of a stimulus and then providing him a corrective feedback. An alternative training method is to provide a participant with a category label simultaneously with the category exemplar. Observational learning has been shown to sometime alter how categories are learned (Ashby, Maddox, & Bohil, 2002; Cincotta & Seger, 2007; Daphna Shohamy et al., 2004), we thus wanted to assess how prototype learning may differ under the two training regimes.

The two variants of the A/nonA task included intentional learning and incidental learning. The mode of learning have been shown to significantly alter both what category representations are acquired (Kemler-Nelson, 1984; Love, 2003; Zeithamova & Maddox, under review) and which neural structures are recruited (Aizenstein et al., 2000; Reber, Gitelman, Parrish, & Mesulam, 2003). We thus wanted to assess whether equating learning mode of the A/nonA task to that of the A/B task would not yield the prototype representation in the A/nonA task identical to that of the A/B task.

The main goal of this chapter was to assess the basic behavioral characteristics, such as prototype effects and difficulty, of the four variants of the prototype learning task when identical category structures and test stimuli are used. Each condition included a training phase and a testing phase. Only the character of a training phase was manipulated, differentially for each task version; the testing phase was identical across conditions. The study was run as four separate experiments, one for each prototype learning variant. Because majority of the procedures were identical in all four experiments, we report them here together, noting the differences as needed.

Methods

Participants

Ninety-seven University of Texas at Austin students participated in the experiment, either as a partial fulfillment of a class requirement or for pay. Twenty-three participants completed the feedback A/B task, 23 participants completed the observational A/B task, 27 participants completed the intentional A/nonA task and 24 completed the incidental A/nonA task. Three participants (2 from the feedback A/B task and 1 from the intentional A/nonA task) reached accuracy significantly below chance (less than .36 proportion correct, or 15 or fewer correct responses out of 42, $p < .05$, binomial test, two-sided). As data from these participants also constituted outlying values on the accuracy and several other performance characterizing measures (more than 2.5 standard deviations from the group means), we excluded them from further analysis. The exclusion did not change pattern of results in any of the reported analyses.

Stimuli and apparatus

The stimuli were cartoon animals that varied along 10 binary dimensions, such as body shape (round or parallelogram), head position (facing forward or downward), tail shape (curled or straight), etc (Figure 5.3), adapted from a prototype learning study of Bozoki and colleagues (Bozoki, Grossman, & Smith, 2006). We chose high-dimensional stimuli to discourage rule-based or exemplar memorization strategies (Minda & Smith, 2001). Categories were based on family resemblance, a characteristics typical of natural categories (Rosch & Mervis, 1975; Wittgenstein, 1953). For each participant, one stimulus served as the category A prototype with all 10 of its feature values being referred to as prototypical features. All other stimuli can be defined relative to the prototype and can differ on 1 – 10 of the prototypical feature values. The stimulus with all 10 non-prototypical features is the B prototype (in the A/B task) and the anti-prototype (in the A/nonA task). The number of non-prototypical features in each stimulus determines its distance from the prototype (see Figure 5.1.). Category A stimuli were defined as those with a distance of 0 – 4 from the A prototype and category B (or non-A) stimuli were defined as those with a distance of 6 – 10 from the A prototype. Stimuli equidistant from the two prototypes were excluded from the study. Binary value dimensions allow to naturally equate category B members with non-A members, as the fewer features a stimulus shares with the category A prototype (the farther it is from A), the more features is shares with the category B prototype (the closer it is to B). Identical stimuli were thus used in the test phase for both the A/nonA and the A/B tasks.

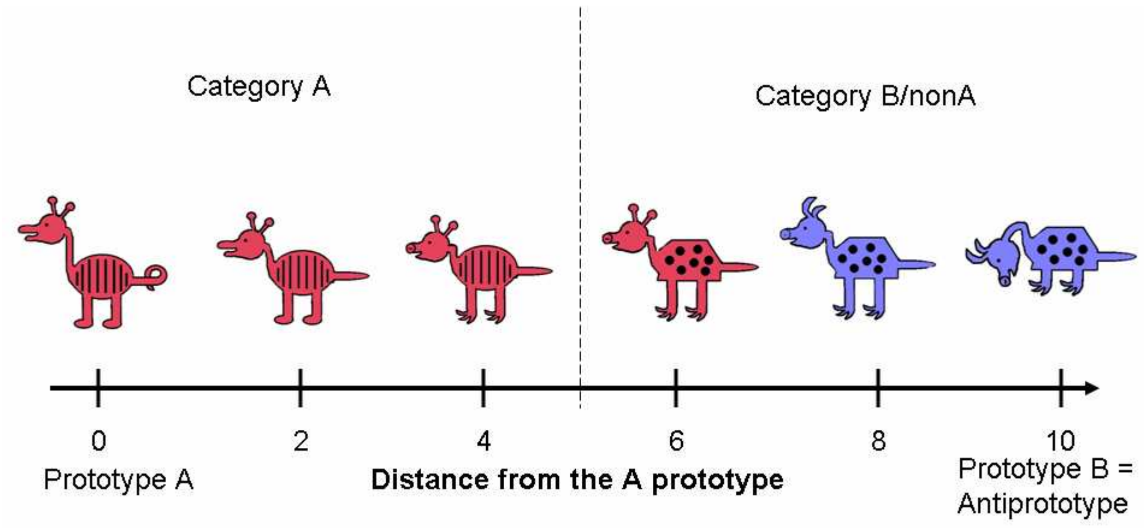

FIGURE 5.3. EXAMPLE STIMULI FROM THE PROTOTYPE TASK STIMULUS SET USED IN THE EXPERIMENT. The left most stimulus represents the prototype of category A, stimuli to the right from the prototype represent examples of stimuli with increasing distances from the A prototype. The right most stimulus represents a category B prototype (or anti-prototype). Stimuli having a distance 0 to 4 from the prototype A were considered category A members, stimuli at the distance 6 to 10 were considered category B (non-A) members.

Procedure

Training phase

Prototype learning variants. Each participant completed 20 training trials of the prototype learning task, presented successively one by one in a random order. In the feedback A/B task, participants were trained to categorize category A stimuli from category B stimuli via corrective feedback. On each trial, 2 seconds after stimulus onset, the participant was prompted to give an A or B response. After each response, the participant was informed whether they were correct or wrong. In the observational A/B task, participants were trained to categorize category A stimuli from category B stimuli by observing the stimuli together with their category label. On each trial, a stimulus was presented and its category label was displayed underneath. After 2 seconds, the participant was prompted to press any key to advance to the next stimuli. In both A/B tasks, 10 A stimuli and 10 B stimuli were presented; within each category, 2 training stimuli differed from the category prototype on 1 feature, 3 differed on 2 features, 3 differed on 3 features and 2 differed on 4 features. Across all 10 stimuli within each category, the category typical features were presented 7 or 8 times and the opposite category typical features were presented 2 or 3 times. Neither prototype was presented. In the intentional A/nonA task, participants were informed that they would be shown examples of category A members and that they would later need to discriminate members of a category (A) from nonmembers (nonA). In the incidental A/nonA, participants were informed that they will be shown some images and later be tested on how well they remember them. After the presentation was completed, participants were informed that all of the stimuli they saw were members of one category and they now need to discriminate members from nonmembers. In both A/nonA tasks, participants were shown stimuli from category A only. The stimuli were passively viewed one by one for a minimum of 2 seconds, after which a prompt asked a participant to press any button to proceed to a next stimulus. There were 5 training stimuli that differed from the A prototype on one feature, 5 differed on two features, 5 differed on three features and 5 differed on four features. Across all 20 stimuli, the prototypical value on each dimension was presented 15 times and the non-prototypical value on each dimension was presented 5 times.

Testing phase

The testing phase was identical in all tasks, with only the label of the second category (B or nonA) differing between the tasks. Participants were presented with 42 stimuli, one at a time. The stimuli included both prototypes and five stimuli selected from each distance from the prototype (except distance 5 - ambiguous stimuli). None of the stimuli were previously used in the training phase. Each trial started with a 500 ms fixation cross, followed by a stimulus. The stimulus was presented until the participant indicated the category membership of the stimulus by pressing one of two buttons. No feedback was provided and inter-trial interval was 1 second.

Results

Proportion of correct responses

First, we compared difficulty of different variants of the prototype learning task. The correct response was to indicate category A to stimuli with distance 0 to 4 features from the A prototype and to indicate category B to stimuli with distance 6 to 10 features from the A prototype. Although this definition is somewhat arbitrary for the A/nonA task (as one may argue that higher distances from the A prototype simply represent higher distortions from the A prototype, or peripheral members of category A), participants were a priori informed that there would be equal number of exemplar who do and who do not belong to the A category and the pattern of their responses indicates that they did adopt this definition. The results are presented in Figure 5.4.

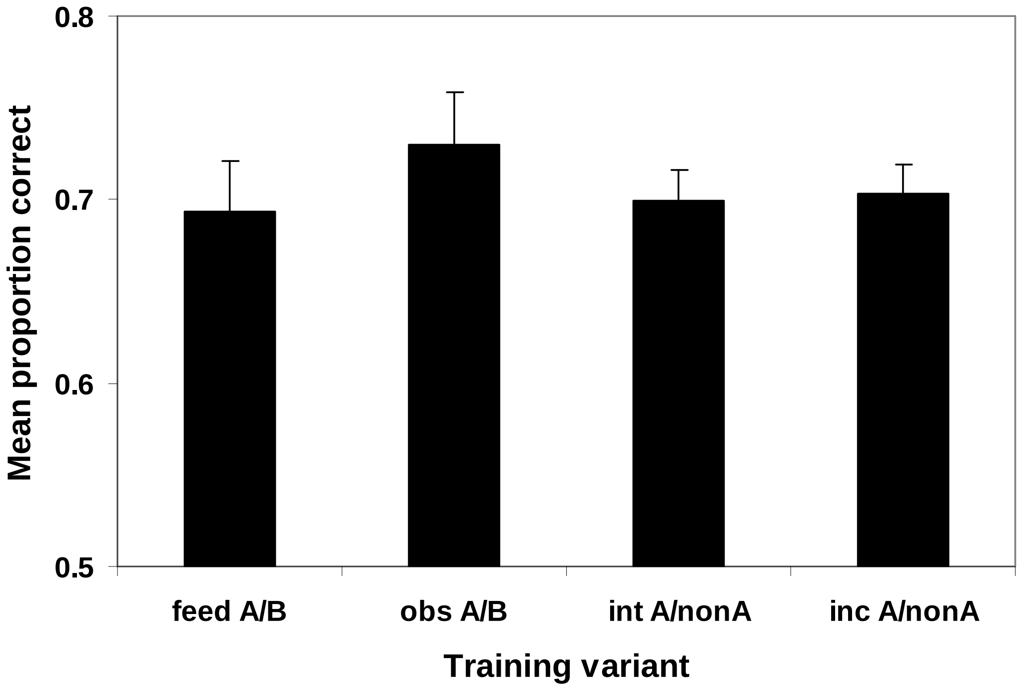

FIGURE 5.4. MEAN ACCURACY IN THE FOUR PROTOTYPE LEARNING VARIANTS. feed A/B: Feedback training A/B task; obs A/B: Observational training A/B task; int A/nonA: Intentional learning A/nonA task; inc A/nonA: Incidental A/nonA learning. The error bars represent the standard error of mean.

As Figure 5.4 suggests, control accuracy was comparable across the four tasks, about .70 proportion correct in feedback A/B, intentional A/nonA and incidental A/nonA training tasks, and .73 proportion correct in observational A/B training task.[15] The equivalent difficulty of the feedback A/B task and the intentional A/nonA task are especially encouraging with respect to our goal of developing a version of the A/B task and the A/nonA task that use the same learning more and testing stimuli and yield comparable performance.

[15] Because the data from different tasks were not collected in one experiment, it is not appropriate to treat them as different levels of the "task" factor.

Category endorsement as a function of the distance from the prototype

A characteristic feature of categorization based on prototype representation is that category membership is graded rather than all-or-none (McCloskey & Glucksberg, 1978). We thus calculated proportion endorsement into category A as a function of the distance from the prototype A). The mean endorsements for the four training variants are presented in Figure 5.5.

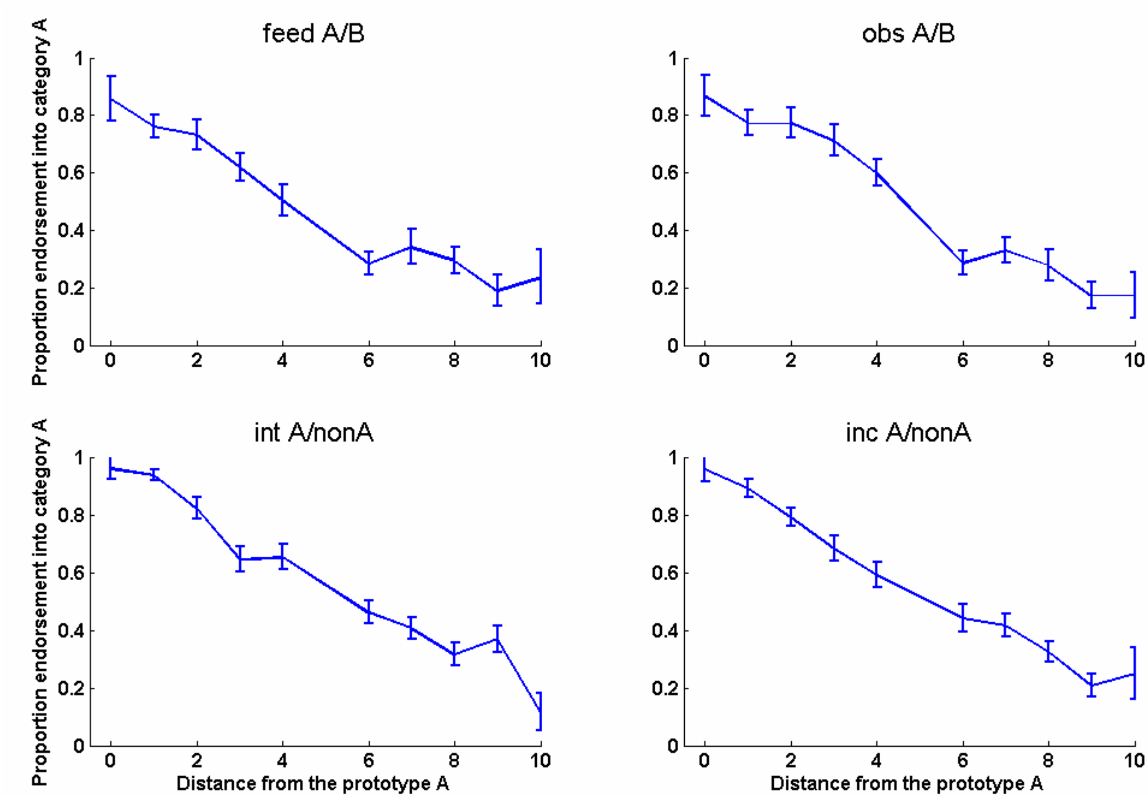

FIGURE 5.5. MEAN PROPORTION ENDORSEMENTS INTO CATEGORY A AS A FUNCTION OF THE NUMBER OF FEATURES SHARED WITH THE A PROTOTYPE.

The first thing to note on Figure 5.5 is that the proportions endorsement show a rather nice linear relationship with the distance from the A prototype. We verified that this relationship exists on individual participant's basis and is not a by-product of averaging across subject by fitting linear trends into individual endorsements and calculating correlation coefficient between the distance from the prototype and the proportion endorsement.[16] We found that 72 out of 94 participants across tasks showed significant linear trend, with median correlation being r = .825. The parameters of the linear fits for each experiment and each condition are presented in Table 5.1.

	slope	intercept	Proportion fit
feed A/B	0.073 (0.010)	0.102 (0.048)	0.62
obs A/B	0.083 (0.011)	0.078 (0.053)	0.70
int A/nonA	0.074 (0.006)	0.204 (0.040)	0.70
inc A/nonA	0.080 (0.005)	0.142 (0.032)	0.75

TABLE 5.1. LINEAR FIT PARAMETERS (STANDARD ERRORS) FOR THE FOUR TRAINING TYPES. Proportion fit denotes proportion of participants that showed significant linear trend.

Accuracy and reaction time as a function of the distance from the boundary

Performance in prototype learning tasks is characterized by advantage for more prototypical over less prototypical stimuli in both accuracy and speed of classification. Figure 5.6 shows the relationship between distance from the category bound and accuracy (left panel) and reaction time (right panel) for the four tasks. In all tasks, proportion correct increased as the distance from the bound increased (feed A/B $b = 0.061$, $t(20) = 4.99$, $p < .001$; obs A/B $b = 0.049$, $t(22) = 3.74$, $p = .001$; int A/nonA $b = 0.070$, $t(25) = 7.11$, $p < .001$; inc A/nonA $b = 0.090$, $t(23) = 10.42$, $p < .001$). In all tasks except the feedback A/B task, reaction time decreased as the distance from the bound increased (feed A/B $b = -0.011$, $t(20) = 0.34$, $p = .735$; obs A/B $b = -0.213$, $t(22) = 3.29$,

[16] Because 0 shared features (B prototype/anti-prototype) and 10 shared features (A prototype) exemplars were shown to each participant only once each, while all other distances from the A prototype 5 times, the endorsement of prototype and anti-prototype were excluded from the calculation of the endorsement slopes and correlations.

p = .003; int A/nonA b = - 0.172, t(25) = 3.22, p = .004; inc A/nonA b = -0.117, t(23) = 2.75, p = .012).[17]

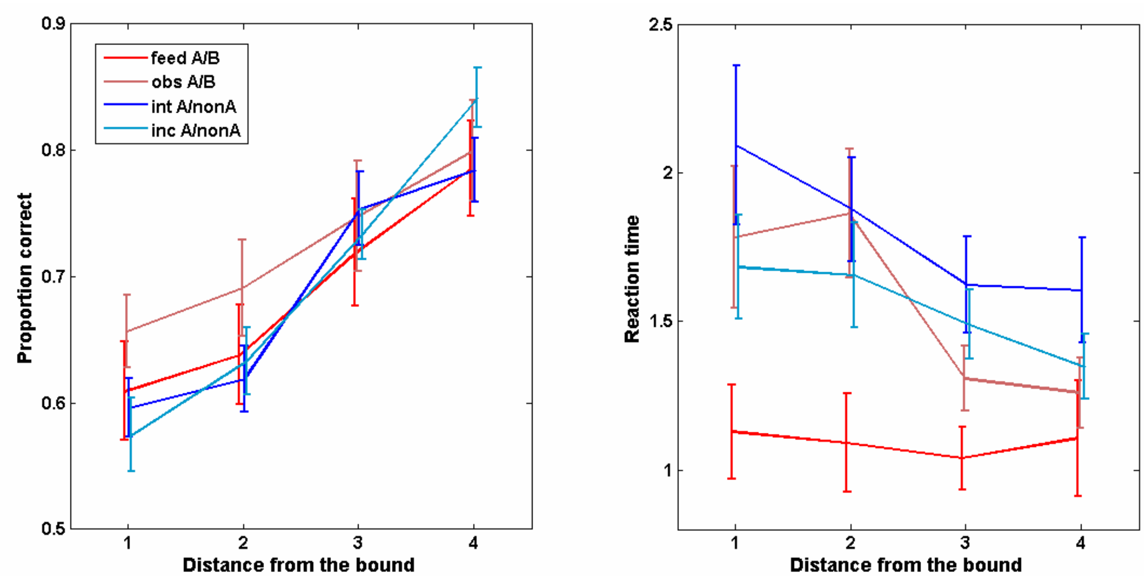

FIGURE 5.6. ACCURACY AND REACTION TIME CHANGES WITH THE DISTANCE FROM THE BOUND. Left: Proportion correct as a function of the distance from the bound. Right: Reaction time as a function of the distance from the bound.

Asymmetric category representation

Figure 5.5 and Table 5.1 suggest that the endorsement intercepts are smaller (and close to zero) in both A/B task variants while they are larger (and above zero) in both A/nonA task variants. Non-zero intercept can be caused both by flatter endorsement slopes (reflecting poorer learning) and/or by an upward shift of the endorsement curve (reflecting category bias). Because category bias would embody the asymmetric representation for the two categories typically found in the A/nonA task (Casale & Ashby, in press; Corneille, Goldstone, Queller, & Potter, 2006; Goldstone, 1996), we wanted to test for it explicitly. We calculated proportion A responses for each participant

[17] We used median reaction time to characterize each participant's reaction time at each distance from the bound. Using mean correct reaction time revealed marginally significant slope in the feedback A/B task (b = -0.064, t(20) = -1.96, p = .065).

and then calculated mean proportion of A responses p(A) in each task. The proportion of A responses was about half in both A/B variants (feedback A/B: $p(A) = .48$, $t(22) = 1.58$, $p = .128$; observational A/B: $p(A) = .49$, $t(11) = 0.77$, $p = .455$), but significantly more than half in both A/nonA variants (intentional A/nonA: $p(A) = .57$, $t(26) = 4.13$, $p < .001$; incidental A/nonA: $p(A) = .55$, $t(23) = 2.70$, $p = .013$). These results suggest that our design can accomplish symmetric representation in the A/B task and asymmetric representation in the A/nonA task even when the actual category structures were equated between the tasks.

Discussion

The results provide three interesting findings about performance patterns in different variants of the prototype learning task. First, all four training methods yielded graded category membership typical for categorization based on prototype representation (McCloskey & Glucksberg, 1978; Rosch, 1975a; Rosch & Mervis, 1975; Rosch & Moore, 1973). The relationship was well described by a linear function between the number of prototypical features and the proportion endorsement. Both accuracy and reaction time also showed a relationship with stimulus prototypicality, with stimuli closer to either prototype being categorized faster and more accurately than stimuli further from the prototypes (close to the boundary).

Second, the average difficulty (proportion correct) was the same irrespective of the training variant. In fact, it was even numerically equivalent in three training variants – feedback A/B, intentional A/nonA and incidental A/nonA (with a small numerical advantage for observational A/B training). This is the first explicit comparison of these training methods. The equivalent difficulty of all four tasks is interesting since it suggests that participants were able to learn just as much about two categories when trained to distinguish between them using exemplars from both, when intentionally trying to learn characteristics of just a single category, and even when incidentally learning characteristics of a single category by being passively exposed to its exemplars. From the methodological perspective, comparable difficulty is important as well as it allows us to draw conclusions about the differences between the task variants without the danger that these differences are attributable to different difficulty of different versions.

Third, both tested variants of the A/B task yielded symmetrical and both tested variants of the A/nonA task yielded asymmetrical category representation, reproducing characteristics traditionally associated with these tasks (Casale & Ashby, in press; Corneille, Goldstone, Queller, & Potter, 2006; Goldstone, 1996). Additional behavioral signature of the A/nonA task was proposed by Casale and Ashby (in press). Casale and Ashby argued that if the A/nonA task is supported by a perceptual learning system, it should exhibit itself by steeper typicality gradient than the A/B task, meaning that accuracy should fall off faster with the distance from a prototype in the A/nonA task than in the A/B task. This follows from the dependence of the perceptual learning system on perceptual similarity. They indeed found steeper accuracy gradients in their A/nonA task (where the noncategorical patterns were random dot formations, unrelated to both the A prototype and each other), however, the performance in their A/nonA task was lower than their A/B task at all levels of exemplar distortions. Alternative explanation thus remained plausible that the steeper gradients are solely driven by greater difficulty of their A/nonA task than their A/B task. Finding accuracy gradient differences here would constitute stronger evidence as both actual stimuli used and overall difficulty of the two tasks were equated. Interestingly, the accuracy slopes observed here were indeed somewhat steeper for both A/nonA variants and flatter for both A/B variants, suggesting that this property of the A/nonA task may be preserved even when category structures of both tasks are equated. Further research is however needed to confirm this observation.

To summarize, Chapter 5 had two intertwined goals. First, we wanted to compare behavioral characteristics of performance in the A/B task and the A/nonA task when confounding differences between them are limited. Second, we wanted to develop an alternative methodology for studying the A/B task and the A/nonA task that would allow their comparison using brain imaging technique. To achieve these goals, we compared learning and categorization performance in two variants of the A/B task and two variants of the A/nonA task using identical stimuli and category structures in all tasks. We found all four training variants to be of comparable difficulty and showing characteristic graded category membership. Importantly, we also found that the new methodology can preserve characteristic profiles and differences between the A/B task and the A/nonA task traditionally associated with these tasks. We conclude that the novel method is a viable alternative to the traditional prototype tasks that provides a new means of comparing the A/B task and the A/nonA task while eliminating confounding differences between them.

Chapter 6. Neural correlates of prototype learning: an fMRI study

Studies of prototype learning have a long tradition in cognitive research (Posner & Keele, 1968, 1970). Despite its importance in everyday cognition, the neural underpinnings of prototype learning are unclear and contradictory findings exist. For example, one critical learning system is episodic memory and examinations of prototype learning in amnesiac patients, with damage to this system, have so far been mixed. Some studies with amnesiacs suggest that prototype learning is intact (Knowlton & Squire, 1993), whereas others suggest a prototype learning deficit (Zaki, Nosofsky, Jessup, & Unversagt, 2003). The picture in the neuroimaging literature is equally inconclusive. Some studies revealed dependence on brain networks associated with episodic memory (Reber, Gitelman, Parrish, & Mesulam, 2003), while others revealed involvement of other networks including those associated with perceptual learning (Reber, Stark, & Squire, 1998a), visuospatial attention and learning (Little & Thulborn, 2005), and visual reasoning (Seger et al., 2000).

In a review of the prototype literature, Ashby and Maddox (2005) suggested that the use of A/nonA tasks in some studies and A/B tasks in other studies explains the contradictory results. They suggested that A/B learning is mediated by medial temporal lobe based explicit episodic memory processes while A/nonA learning is mediated by nondeclarative, perceptual learning processes. Differential involvement of hippocampus in the two tasks can be expected from their cognitive demands. In the A/nonA task, participants are likely to form an isolated representation of a single prototype (Goldstone, 1996) and then compare each test item to this single prototype. If the new stimulus is sufficiently similar to the prototype representation, it will be endorsed to the category; otherwise it will be categorized as a non-member. Changes in representation of categorical items within visual cortex may be used as a basis for successful categorization (Reber, Stark, & Squire, 1998b). In the A/B task, participants are likely to form representations of two distinct categories centered on two prototypes. Each new stimulus is then compared to both of these prototypes and endorsed to the category of the prototype that is closer to the current stimulus. Associations among characteristic features of each category must be formed, together with the category label. Hippocampus has been shown to mediate such learning of arbitrary novel associations (Moss, Mahut, & Zola-Morgan, 1981; Wirth et al., 2003).

Existing literature offers only partial support for the declarative versus nondeclarative memory for the A/B versus A/nonA task. One one hand, impaired prototype learning was reported in amnesia when the A/B task was used (Zaki, Nosofsky, Jessup, & Unversagt, 2003) but spared prototype learning was reported in amnesia when the A/nonA task was used (Bozoki, Grossman, & Smith, 2006; Keri, Kalman, Kelemen, Benedek, & Janka, 2001). The results from neuroimaging studies are less cohesive. Hippocampus involvement has been reported in some studies that employed an A/B task (DeGutis & D'Esposito, 2007; Little, Shin, Sisco, & Thulborn, 2006), but not uninamously (Little & Thulborn, 2005, 2006). Additionally, hippocampal or medial temporal lobe involvement has also been reported in some studies that employed an A/nonA task (Aizenstein et al., 2000; Reber, Gitelman, Parrish, & Mesulam, 2003).

Direct comparison of the results from the existing A/B and A/nonA studies is complicated by a number of methodological differences. First, as outlined above, the category structures differ across tasks. In the A/B task, each category is internally consistent and consists of a collection of exemplars derived from the category prototype. In the A/nonA task, one category is internally consistent, and the other is not. Second, the A/B task involves intentional learning where the participants are instructed to learn the characteristics of the categories based on corrective feedback. The A/nonA task often involves incidental learning where participants passively view category exemplars first, and are assessed on discrimination of categorical from non-categorical exemplars later. Third, different fMRI contrasts have been used in the A/B studies and A/nonA studies. In A/B studies, the BOLD signal was typically contrasted between the prototype task and fixation cross viewing (Little & Thulborn, 2005; Seger et al., 2000), whereas in A/nonA studies, the BOLD signal was contrasted between categorical (A) items and non-categorical (nonA) items, both being a part of the prototype task (Reber, Stark, & Squire, 1998b; Reber, Wong, & Buxton, 2002).

The goal of this study is to examine the neural underpinnings of A/B and A/nonA prototype learning using experimental methodology that holds constant the category structures, learning mode, and fMRI contrast. We will focus on early learning (20 training trials) as most prototype learning studies that we discussed so far used a small number of training trials. We hypothesize that A/B performance is mediated by the medial temporal lobe while A/nonA performance involves nondeclarative learning. Differences in learning mode, category structure and fMRI contrasts between the A/B and A/nonA task could account for a number of discrepancies in the existing literature and must be controlled. Rather than being driven by the specific category structures, we expect that the context of learning – learning contrasting categories or learning characteristics of a single category – can by itself produce the dissociation between declarative and nondeclarative based learning.

Method

Participants

Twenty-seven young adult (age 18-30) volunteers (13 females) participated in the study. Data from 3 participants (1 female) were excluded from analysis due to excessive head motion, leaving 24 participants for analysis. Each participant read and signed informed consent with participation in an fMRI study. Volunteers received $50 compensation for a 2-hour session.

Stimuli

Two sets of stimuli were used in the study. The first set was identical to the set used in Chapter 5 (Figure 5.3). A second set of cartoon animal stimuli with different dimensions was also generated (Figure 6.1), and each prototype learning task was tested with both sets of stimuli. Note that in this study, unlike in a typical A/nonA experiment, all nonA stimuli were internally consistent and constructed from a fixed prototype. Thus the only difference between the A/nonA and A/B tasks was in stimuli presented during training (only A stimuli in the A/nonA task, and A and B stimuli in the A/B task), and the category labels used during the testing phase. Critically, the same stimuli were used in the test phase for both the A/nonA and A/B tasks. Thus, any differences observed in the A/nonA and A/B brain activations cannot be attributed to differences between the structures of nonA category versus B category, nor to any stimulus-specific differences.

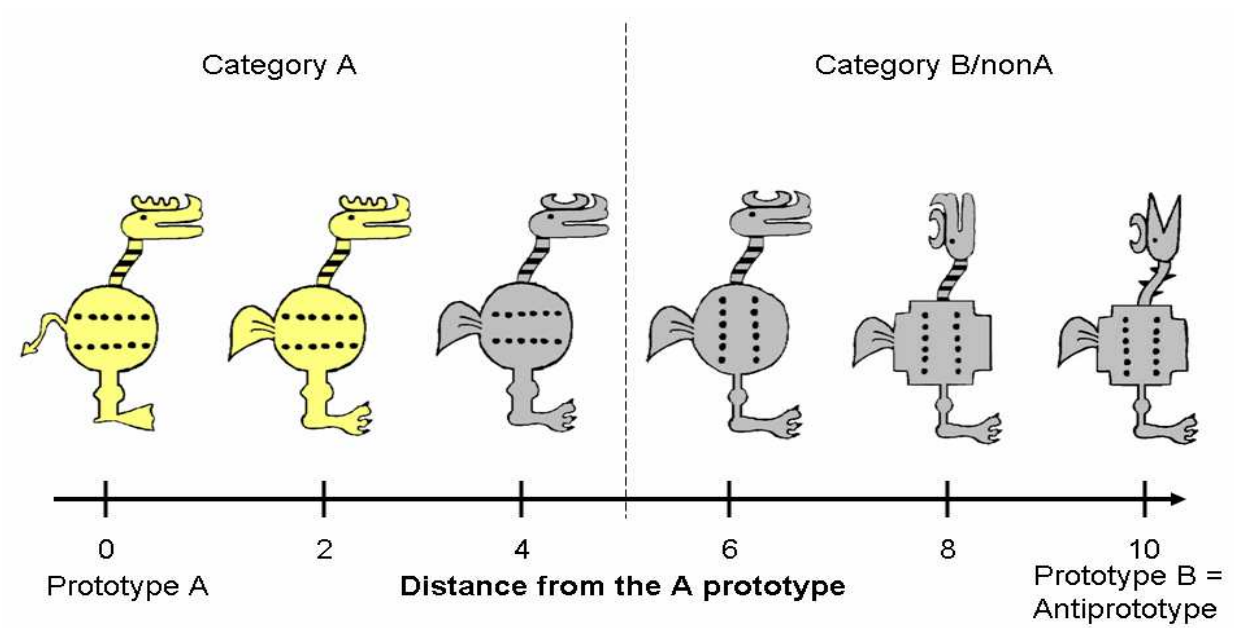

Figure 6.1. Example stimuli from the second stimulus set.

Experimental design

A within subject design was employed. Each participant completed an A/nonA and an A/B run with each stimulus set (4 runs total). Each participant completed two runs of one task, a 10 minute structural scan, and then two runs of the second task. The order of stimulus sets and the order of the tasks were counterbalanced between participants. Each run consisted of a training and test phase. Importantly, although the training phase differed across tasks, the test phase was identical. Functional MRI scans were acquired during the testing phase only.

Training design (not scanned)

The A/B task procedure was identical to the feedback A/B training variant from Chapter 5. Participants were asked to categorize 10 A and 10 B items, using corrective feedback to learn the category labels. The A/nonA task procedure was identical to the intentional A/nonA training variant from Chapter 5. Prior to A/nonA training, participants were informed that they will need to learn to discriminate members of a category (A) from nonmembers (nonA). During A/nonA training, participants were shown 20 stimuli from category A only. The training stimuli were presented in a random order.

Test design (scanned)

The testing phase was identical for both tasks, with only the label of the second category (B versus nonA) differing between the tasks. Participants were presented with 42 stimuli, one at a time, that included both prototypes and five stimuli selected from each distance from the prototype (except distance 5 - ambiguous stimuli). None of the stimuli were previously used in the training phase. An event-related design was utilized to study the neural activity during the testing phase. Four possible orders of A and B stimuli and their onsets including 30% of null time (to interject temporal jitter) were predetermined using “optseq2” program (http://surfer.nmr.mgh.harvard.edu/optseq, Dale, 1999). Each stimulus onset time and order was used in one experimental run. On each trial, a stimulus was presented for a maximum of 3.5 sec, during which time the participant needed to indicate the category membership of the stimulus. No feedback was provided. A fixation cross was presented between each stimulus onset lasting either 0.5, 2.5, or 4.5 seconds.

MRI acquisition, processing and analysis

Functional and structural images were acquired using a 3T GE Signa MRI scanner. Functional images were acquired during the testing phase of each task only, using a multiecho GRAPPA parallel imaging EPI sequence that reduces typical EPI distortions and susceptibility artifacts. Images were collected utilizing whole head coverage with slice orientation to reduce artifacts (approx 20 degrees off the AC-PC plane, TR = 2 sec., 3 shot, TE = 30 msec., 35 axial slices oriented for best whole head coverage, acquisition voxel size = 3.125 X 3.125 X 3 mm with a .3 mm inter-slice gap). The first four EPI volumes were discarded to allow scans to reach equilibrium. Stimuli were viewed through a back projection screen and a mirror mounted on the top of the head coil. Responses were collected with an MR compatible button box that was placed under the right hand.

In addition to collecting EPI images during task performance, one or two high resolution T1 SPGR scans that have been empirically optimized for high contrast between gray matter (GM) and white matter (WM), and GM and cerebrospinal fluid (CSF) were acquired. These images were acquired in the sagital plane using a 1.3 mm slice thickness with 1 square mm in-plane resolution.

Pre-processing and data analysis were conducted using FEAT (FMRI Expert Analysis Tool) Version 5.63, part of FSL (www.fmrib.ox.ac.uk/fsl) software. Pre-processing included motion correction using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002), non-brain removal using BET (S. Smith, 2002), high-pass temporal filtering with a 60 second cut-off, and spatial smoothing with a Gaussian kernel of 5 mm FWHM. Data from each run of each participant were analyzed separately at a first level of analysis. Each category stimulus time onset was convolved with a canonical hemodynamic response function and was entered as a predictor into a general linear model to estimate β-weights together with their temporal derivatives. Data from all four runs from each participant were combined at a second level using a fixed effects analysis. Combined data from each participant were then subjected to the third level random effects analysis using OLS. For all analyses, individual voxels were considered active when reaching Z > 2.3. Whole brain cluster-size threshold was set at $p < .05$ (Worsley, 2001). Additionally, we defined two regions of interest (ROI), medial temporal lobe (MTL) and striatum. We were especially interested in area MTL because its involvement in prototype learning has been controversial and in striatum because it has been implicated in other kinds of category learning and is thought to operate complementary to area MTL (Poldrack et al., 2001; Poldrack & Packard, 2003). The MTL ROI consisted of FSL Harvard-Oxford atlas defined right and left hippocampus and right and left parahippocampus, the striatum ROI consisted of FSL Harvard-Oxford atlas defined right and left putamen and right and left caudate. Activation in each ROI was assessed using a small volume correction at $p < .05$ based on Monte Carlo simulation using AFNI, accounting for both smoothness of the data and the shape and size of each ROI. The simulation determined a minimal required cluster size of 33 voxels for the MTL ROI and 30 voxels for the striatum ROI.

Results

Behavioral performance

For the main behavioral and fMRI analyses, test phase data were pooled across the two runs (the two stimulus sets) of each task.[18] Endorsement functions – observed probabilities of responding A at each distance from the prototype A – are presented in Figure 6.2.[19] To ensure that the linear trend is not a by-product of averaging across participants (Maddox, 1999), we calculated endorsement slopes for each participant separately, excluding the distance 0 (the prototype A) and distance 10 (the prototype

---

[18] There were no differences between accuracies achieved on the two stimulus sets (A/B: .68 vs .71, $t(23)= 1.186$, $p=.248$; A/nonA: .67 vs .68, $t(23)=0.441$, $p=.664$).

[19] Three different participants reached accuracy significantly below change (proportion correct less than .36 (15 or less correct responses out of 42), $p < .05$, binomial test, two-sided) in one of their four runs. For these runs, we assumed that the participant actually learned to distinguish between the two categories, but confused the labels during the test, and adjusted the scoring accordingly. We also repeated all behavioral analyses with (1) the original data and with (2) the three runs excluded; the pattern of results was not affected in either case.

B/anti-prototype) as they consisted of only one data point. For both tasks, the proportion of category A endorsements decreased linearly with the distance from the prototype A, with mean slopes similar in both tasks (A/B $b = 0.059$, $se = 0.008$, $p < .001$, A/nonA $b = 0.054$, $se = 0.007$, $p < .001$; slope difference: $t(23) = 0.386$, $p = .703$), and with no bias for either response in the A/B task (proportion of A response = .494, $t(23) = 0.63$, $p = .532$), but with a bias towards A response in the A/nonA task (proportion of A response = .580, $t(23) = 3.82$, $p < .001$). In both tasks, categorization accuracy increased linearly as a function of a stimulus distance from the category boundary (A/B: $b=0.058$, $se=0.010$, $p < .001$, A/nonA: $b=0.060$, $se=0.009$, $p<.001$; slope difference: $t(23) = 0.232$, $p = .818$) and reaction times decreased linearly as a function of a stimulus distance from the category boundary (A/B: $b=0.029$, $se = 0.011$, $p=.020$; A/nonA: $b=0.028$, $se=0.013$, $p=.045$; slope difference: $t(23) = 0.082$, $p = .935$). Importantly, there was no difference between overall accuracy in the A/B (mean = .694, se = .018) and A/nonA tasks (mean = .673, se = .020; $t(23) = -0.644$, $p = .526$). Interestingly, A/B and A/nonA accuracy rates were moderately negatively correlated ($r = -.362$, $p = .082$), suggesting that distinct cognitive processes may underlie performance in the two tasks. Unlike accuracy, mean reaction times differed between the two tasks by approximately 0.2 s (A/B: mean = 1.343 sec, se = .095; A/nonA: mean = 1.545 sec, se = 0.099; $t(23) = 3.566$, $p = .002$) and were positively correlated within subjects ($r=.831$, $p<.001$).

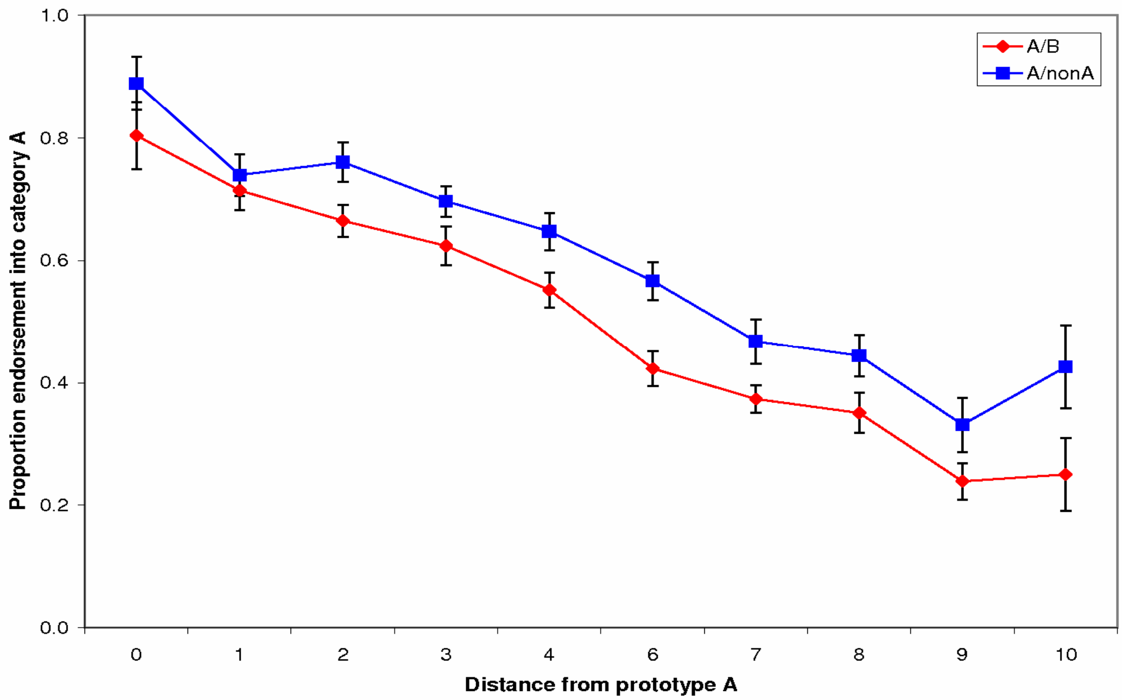

Figure 6.2. Endorsement into category A as a function of a stimulus distance from prototype A.

Common neural regions

First, we identified regions that showed common activation or common deactivation in both the A/B task and the A/nonA task compared to the fixation baseline, by creating a conjunction z-map using the minimum of the two tasks' z-maps. A network of regions in which both tasks showed significantly greater activation compared to the fixation baseline (see Figure 6.3, Table 6.1) included areas involved in visual perception and object identification (occipital and fusiform areas, Figure 6.3.a,b), areas associated with decision making and response generation (inferior frontal cortex and precentral gyrus, Figure 6.3.b), as well as bilateral posterior hippocampus (Figure 6.3.c) and bilateral striatum (Figure 6.3.d). A number of regions in which both tasks showed deactivation compared to the fixation baseline (see Figure 6.4., Table 6.2.) were identified as well, consisting primarily of typical default-mode network regions

(Laurienti, 2004; Mason et al., 2007; Raichle et al., 2001): inferior parietal cortices, posterior cingulate, medial temporal cortices and medial frontal cortices.

Brain region	Size	Max Z	x	y	z
Whole brain cluster corrected (p < .05)					
L lateral occipital (BA 19)	4719	6.93	-36	-86	-4
R lateral occipital (BA 19)	3387	6.75	42	-66	-12
Calcarine (BA 17)	807	3.59	10	-72	6
L Postcentral (BA 3/40)	3705	5.71	-44	-26	48
R Inferior Parietal (BA 7/40)	3513	5.69	36	-54	46
R Fusiform/Inferior temporal (BA 37)	1737	6.83	40	-54	-20
L Fusiform (BA 37)	896	6.73	-38	-64	-20
L Inferior frontal (BA 44/48)	2333	6.1	-48	6	28
Medial frontal (BA 24/32)	1927	5.79	-4	8	46
R Inferior frontal (BA 44)	1740	5.61	52	10	24
R Middle frontal (BA 6)	489	4.08	30	-4	46
Small volume corrected (p < .05)					
R Hippocampus	123	5.15	20	-30	-4
L Hippocampus	84	4.56	-20	-30	-8
R Striatum	130	4.35	16	16	-2
L Striatum	61	3.23	-20	10	-4

TABLE 6.1. REGIONS COMMONLY ACTIVATED IN BOTH A/B AND A/NONA TASK. L = left, R = right, BA = Broadman area, Max = maximum. Size given in voxels.

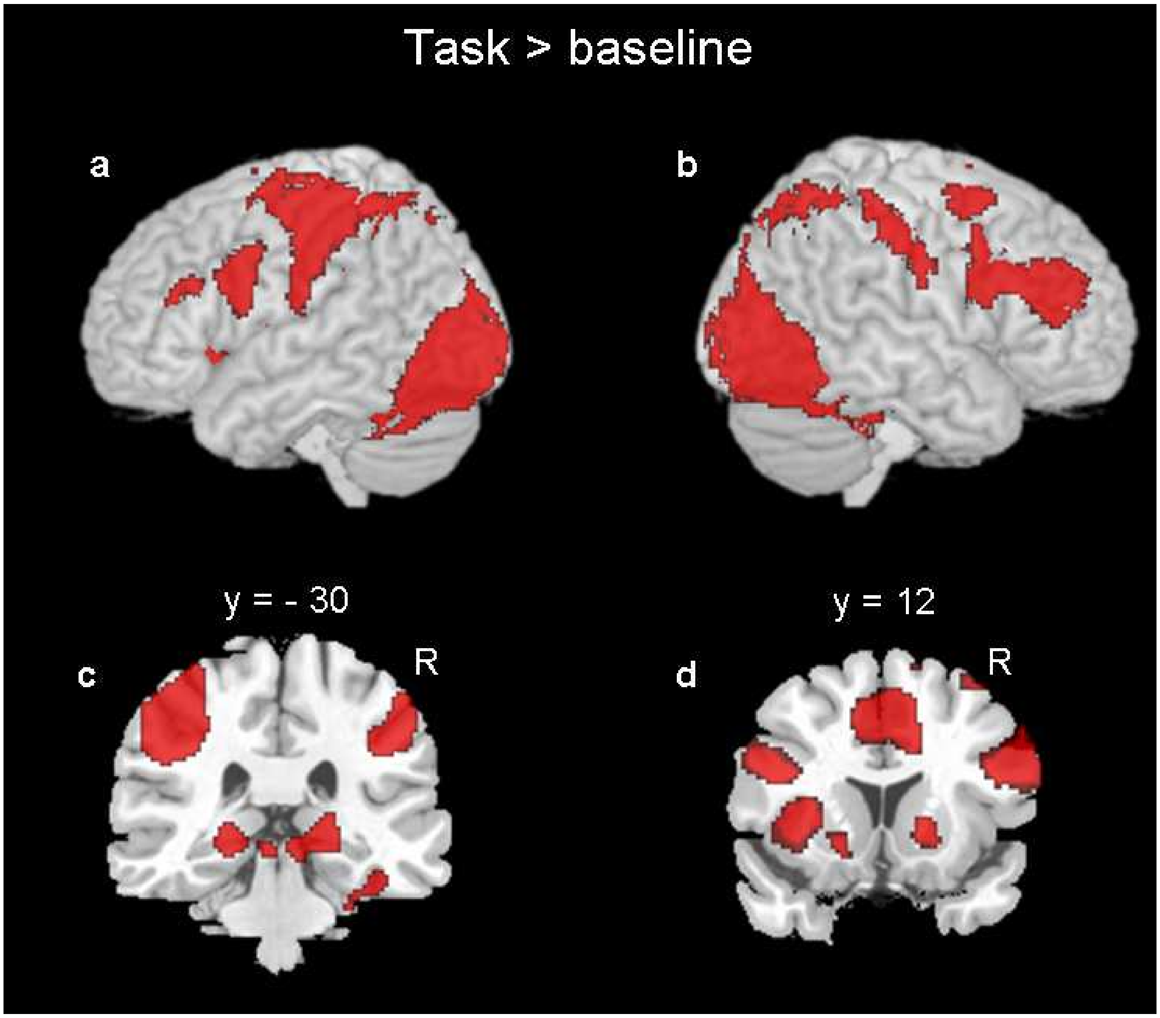

FIGURE 6.3. COMMONLY ACTIVATED REGIONS FROM BOTH TASKS VERSUS BASELINE. a, b: Whole brain 3D rendering with cortical activation overlay. a. Left hemisphere. b. Right hemisphere. c, d: Coronal slices with activations overlays. c. Bilateral hippocampus. d. Bilateral striatum and medial frontal cortex. Activation maps were overlaid upon a canonical brain using MRIcro software (www.mricro.com)

Brain region	Size	Max Z	x	y	z
Whole brain cluster corrected (p < .05)					
Posterior cingulate (BA 23)	4494	5.49	-10	-42	40
L Angular gyrus (BA 39)	2008	5.79	-56	-62	22
R Supramarginal (BA 40)	1880	5.3	58	-48	24
R Middle temporal (BA 20/21)	4887	5.45	52	-4	-32
L Middle temporal (BA 20/21)	4878	5.5	-62	-48	-2
Superior frontal (BA 9)	8151	5.56	-28	26	46
L Orbito frontal (BA 38/11)	1002	4.75	-52	24	-10
R Orbito frontal (BA 47/11)	686	4.6	50	34	-10
L Medial temporal	804	4.93	-24	-16	-22
R Medial temporal	636	4.49	22	-12	-30

TABLE 6.2. REGIONS COMMONLY DEACTIVATED IN BOTH A/B AND A/NONA TASK. L = left, R = right, BA = Broadman area, Max = maximum. Size given in voxels.

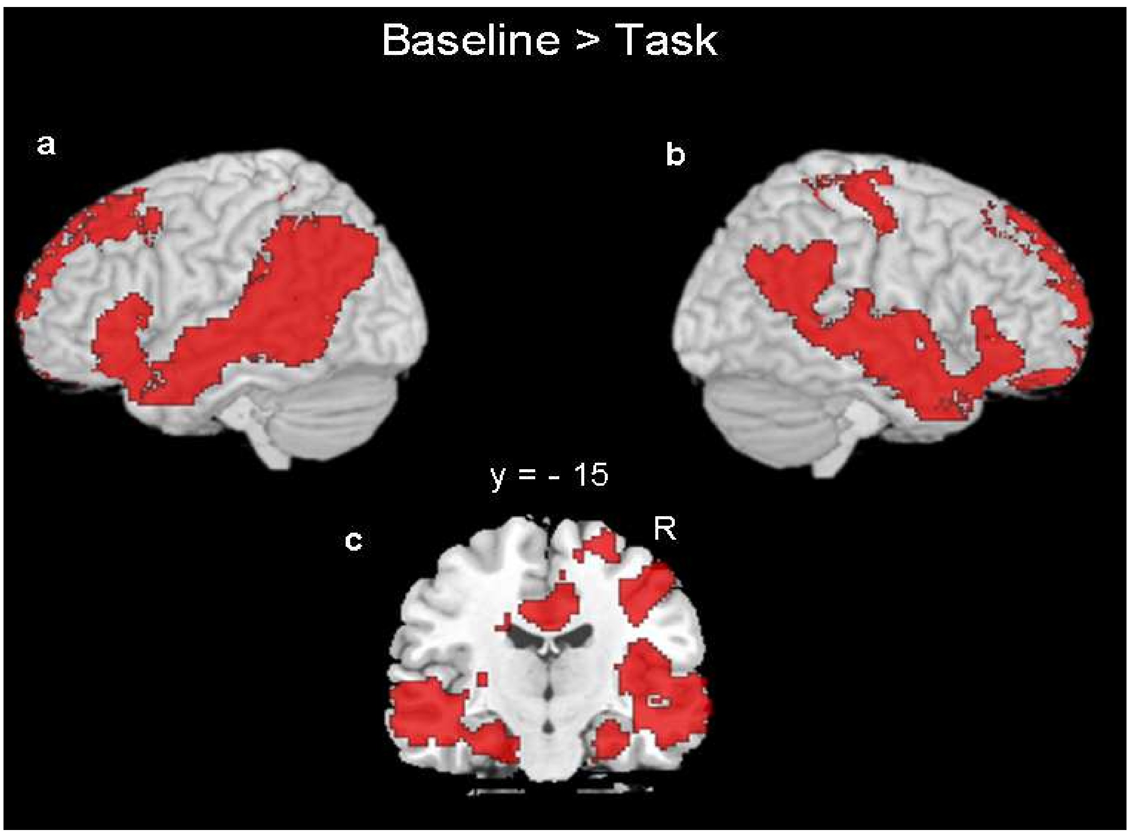

FIGURE 6.4. COMMONLY DEACTIVATED REGIONS FROM BOTH TASKS VERSUS BASELINE. a, b: Whole brain 3D rendering with cortical activation overlay. a. Left hemisphere. b. Right hemisphere. c, Coronal slice showing MTL regions. Activation maps were overlaid upon a canonical brain using MRIcro software (www.mricro.com)

Distinct neural regions

The primary goal of this research was to directly compare activity during the A/B task and the A/nonA task, controlling the stimuli and learning mode. A number of regions exhibited increased activity in one task compared to the other. The list of identified regions is provided in Table 6.3., contrast activation maps are provided in Figure 6.5.

Brain region	Size	Max Z	x	y	z
A/B > A/nonA (Whole brain cluster corrected)					
R Inferior parietal (BA 40)	746	3.79	58	-38	46
L Orbito frontal (BA 47/11)	381	3.9	-36	54	-16
A/B > A/nonA (Small volume corrected)					
L Parahippocampus (BA 36)	53	3.13	-20	-6	-30
A/nonA > A/B (Whole brain cluster corrected)					
L Inf Lateral Occipital (BA 18)	989	4.88	-20	-94	0
R Inf Lateral Occipital (BA 19)	776	4.37	38	-82	-2
L Sup Parietal (BA 7)	478	4.24	-20	-70	36
R Sup Parietal (BA 7)	451	3.74	22	-64	48
A/nonA > A/B (Small volume corrected)					
R Putamen	41	3.47	20	10	-8
R Caudate head	33	3.24	10	10	-2
L Caudate body	32	4.27	-10	6	18

TABLE 6.3. REGIONS FROM WHOLE BRAIN AND REGION OF INTEREST (SMALL VOLUME CORRECTED) ANALYSIS THAT ACTIVATED DIFFERENTIALLY DURING THE A/B TASK AND THE A/NONA TASK. L = left, R = right, Inf = inferior, Sup = superior, BA = Broadman area, Max = maximum. Size given in voxels.

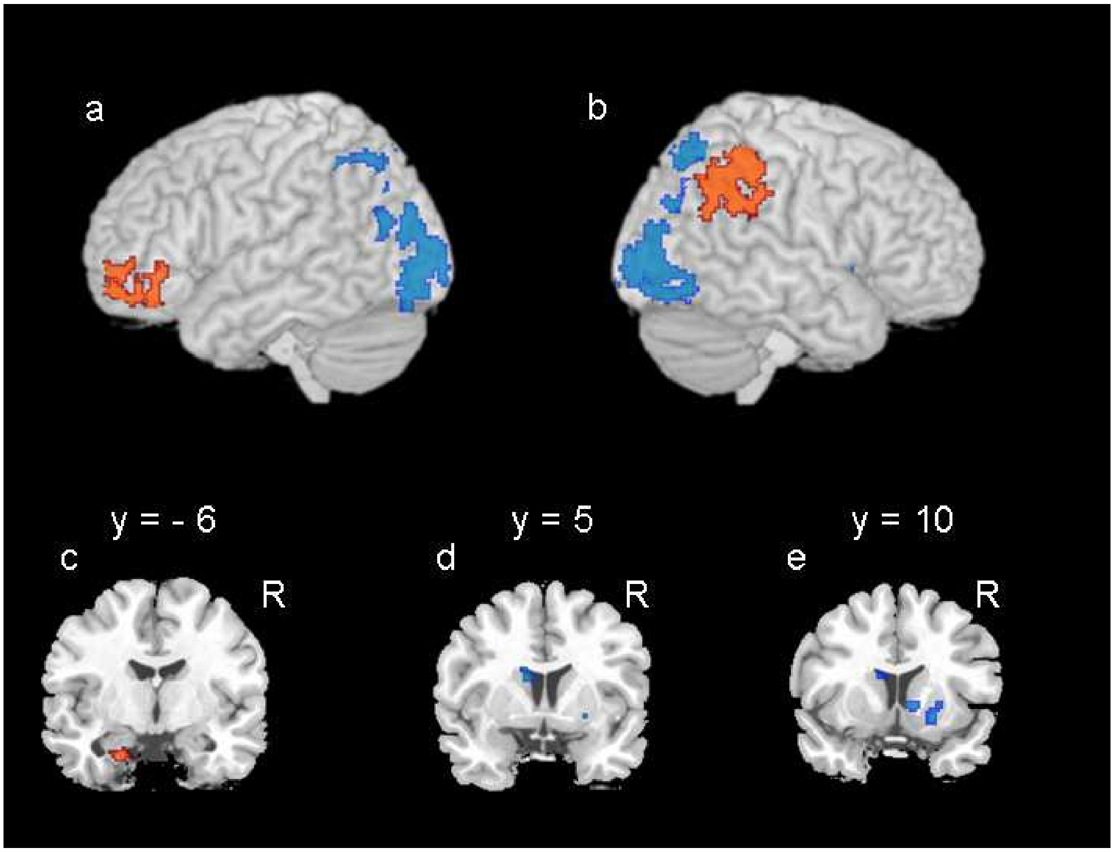

FIGURE 6.5. REGIONS FROM DIRECT CONTRAST OF A/B TASK VERSUS A/NONA TASK. In red, A/B > A/nonA; in blue A/nonA > A/B. a, b: Whole-brain cluster corrected contrasts overlaid on a 3D rendering of a canonical brain. a. Left hemisphere. b. Right hemisphere. c, d, e: Coronal sections illustrating small volume corrected contrast maps in regions-of-interest. c. Left parahippocampus. d. Left caudate body. e. Right putamen and right caudate head. Activation maps were overlaid upon a canonical brain using MRIcro software (www.mricro.com).

The direct contrast revealed that the A/B task involves to a larger degree areas that have been implicated in explicit episodic memory, including frontal and parietal cortices and parahippocampus (Figure 6.5, red overlay). By contrast, regions that demonstrated greater activity in the A/nonA task than the A/B task included those previously implicated in perceptual learning, including posterior cortices and striatum (Figure 6.5, blue overlay).[20]

Neural regions predictive of accuracy

To identify brain areas predictive of successful categorization, we compared activity evoked during correct categorization trials with activity evoked during incorrect categorization trials, separately for each task. Identified regions that exhibited greater activation during correct than incorrect trials are listed in Table 6.4 and are presented in Figures 6.6 and 6.7. No region exhibited greater activation for incorrect than correct trials in either task. Regions that were predictive of correct categorization during the A/B task trials included bilateral middle temporal cortices, posterior cingulate cortex and orbito frontal cortex, as well as bilateral medial temporal lobe spanning parts of both parahippocampus and hippocampus. Only two regions were predictive of correct categorization during the A/nonA task, left putamen and right anterior hippocampus. The relative location of the right hippocampal region identified in the A/nonA task and the right MTL region identified in the A/B task is presented in Figure 6.8. The A/nonA region was located anterior to the A/B region and there was minimal overlap between the regions (3 voxels).

[20] Because reaction times were not perfectly equated in the two tasks, it is possible that some of the regions identified in the A/nonA > A/B contrast may reflect longer processing times in the A/nonA task than in the A/B task. Although we cannot rule this possibility based on the current data, two pieces of evidence suggest that it is not the case. First, most regions identified in the A/nonA > A/B contrast were either a priori expected (e.g. posterior cortices based on the perceptual learning theory) or were identified also in another contrast (striatum in the correct > incorrect contrast). Second, adding the reaction time differences as a covariate at the group level analysis did not abolish the activation differences.

Brain region	Size	Max Z	x	y	z
A/B task (Whole brain cluster corrected)					
R Middle temporal (BA 21/22)	373	3.64	62	-14	-16
L Middle temporal (BA 21/22)	263	3.57	-58	-36	4
Posterior cingulate/ Precuneus (BA 23)	1576	4.58	-6	-52	10
Orbito frontal (BA 10/11)	1430	4.1	2	62	-14
A/B task (Small volume corrected)					
R Medial temporal (BA 20)	238	4.18	30	-22	-16
L Medial temporal (BA 20)	201	3.69	-32	-22	-14
A/nonA task (Small volume corrected)					
L Putamen	177	4.25	-32	-10	-2
R Anterior Hippocampus	38	3.8	28	-10	-22

TABLE 6.4. REGIONS THAT EXHIBITED GREATER ACTIVATION DURING CORRECT THAN INCORRECT TRIALS. Regions identified for each task separately.

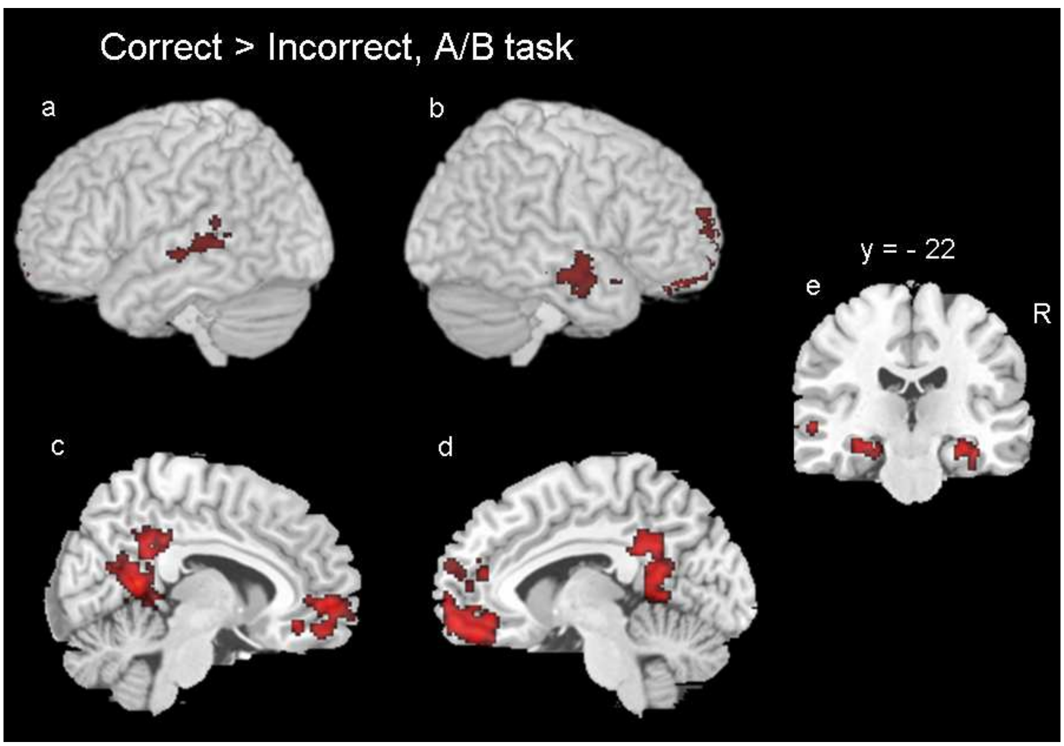

FIGURE 6.6. REGIONS ASSOCIATED WITH SUCCESSFUL CATEGORIZATION DURING THE A/B TASK. a,b: Lateral view of the left and right hemisphere 3D rendering with activation overlay. c,d: Medial view of the left and right hemisphere. e: Coronal section showing medial temporal lobe activation.

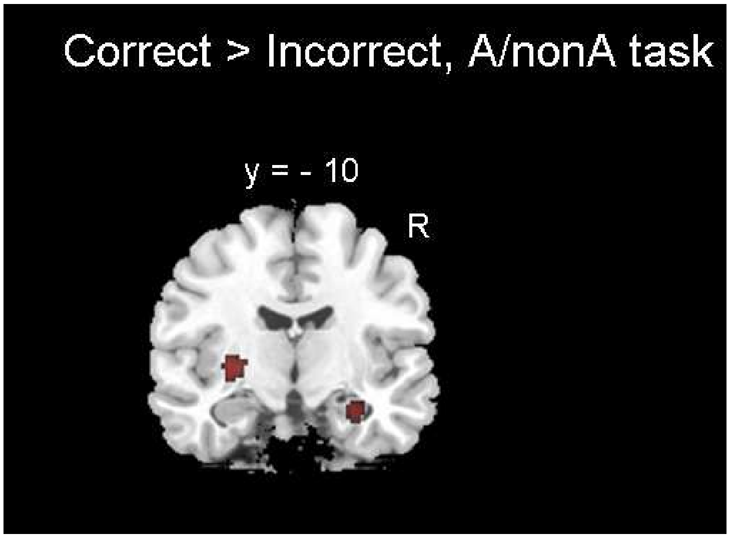

FIGURE 6.7. REGIONS ASSOCIATED WITH SUCCESSFUL CATEGORIZATION DURING THE A/NONA TASK. Coronal section featuring left putamen and right hippocampal activation.

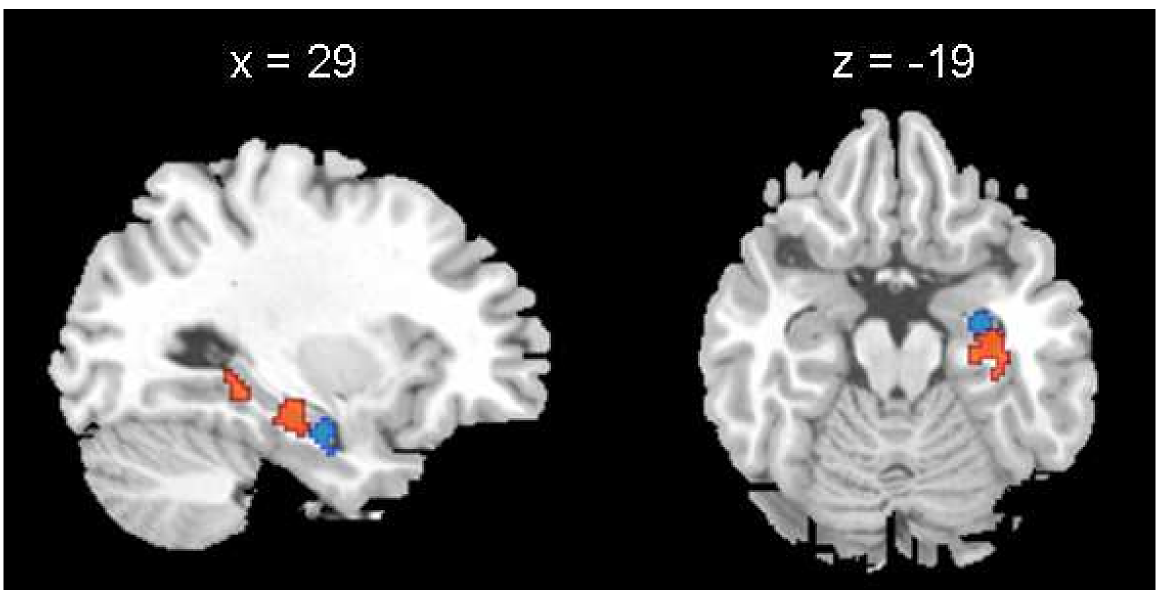

FIGURE 6.8. COMPARISON OF MTL REGIONS IMPLICATED IN THE A/B TASK AND THE A/NONA TASK. Sagital and horizontal section illustrating relative location of the regions of MTL that showed greater activation in correct than incorrect trials during the A/B task (red) and the A/nonA task (blue).

Discussion

Prototype learning is ubiquitous in everyday cognition. We hypothesized that prototype learning is not mediated by a single neural system, but rather that the system relevant to prototype learning depends critically upon the circumstances of learning – whether the task involves learning to discriminate a single category from other stimuli, or two categories from each other. Holding constant the learning mode, structure of the categories, and fMRI contrasts, the results presented here suggest that A/nonA and A/B prototype learning are supported by dissociable cognitive and neural processes. First, we found a negative correlation between A/nonA and A/B task behavioral performance even when learning was comparable in both tasks. Lack of performance correlation between two tasks is often used as an indicator the two tasks rely on different cognitive processes (Conway & Engle, 1996; Kane & Engle, 2003; Shah & Miyake, 1996). Second, we found dissociable neural systems supporting the two tasks using functional MRI. Although there were a number of regions commonly activated or deactivated during both the A/nonA task and the A/B task compared to a fixation baseline, a direct contrast of stimulus evoked activity during the two tasks revealed several regions that were preferentially active during one task versus another,. Most notably, the A/B task recruited to a larger degree parahippocampus, inferior parietal and orbito-frontal cortex, while the A/nonA task recruited to a larger degree lateral occipital and posterior parietal cortices, and striatum. We also identified regions that were predictive of correct categorization by contrasting neural activity during correct and incorrect trials. We confirmed the role of the medial temporal lobes and orbito-frontal cortex in the A/B task and the role of striatum in the A/nonA task.

These findings are generally consistent with findings from several prototype learning studies in amnestic patients. Knowlton and Squire (1993) found intact A/nonA prototype learning in patients with MTL lesion-based amnesia; Bozoki et al. (Bozoki, Grossman, & Smith, 2006) and Keri et al. (Keri, Kalman, Kelemen, Benedek, & Janka, 2001) found relatively spared A/nonA learning in patients with Alzheimer's disease; and Zaki et al. (Zaki, Nosofsky, Jessup, & Unversagt, 2003) found impaired learning in the

A/B task, but not in the A/nonA task, in two groups of amnesic patients. The involvement of the MTL in the A/B task has also been implicated in a few neuroimaging studies. Although some A/B fMRI studies did not report hippocampal activation when contrasting the A/B task versus baseline (Little & Thulborn, 2005; Seger et al., 2000), others did when using more detailed contrasts (low versus high distortions: DeGutis & D'Esposito, 2007; early versus late learning: Little, Shin, Sisco, & Thulborn, 2006).

The role of the medial temporal lobe in prototype learning

Although largely confirming the notion of Ashby and Maddox (2005) that the A/B task may rely on declarative learning supported by the structures of the medial temporal lobe whereas the A/nonA task relies on nondeclarative learning, the declarative versus nondeclarative distinction seems to constitute an incomplete description of the two tasks. Using the contrast of correct versus incorrect trials, we found evidence of hippocampus involvement in the A/nonA task as well. Others also found evidence for MTL involvement in the intentional A/nonA task by contrasting categorical (A) and noncategorical (non-A) stimulus evoked activity (Aizenstein et al., 2000; Reber, Gitelman, Parrish, & Mesulam, 2003). So, how can these seemingly paradoxical results be reconciled? Below, we discuss a number of (mutually nonexclusive) possible accounts.

First, a simple explanation may be that although MTL can be recruited during the A/nonA task, it is not critically necessary for the task performance. It is unlikely that normal healthy participants would not use all resources to perform a task; from this perspective, neuroimaging and neuropsychological data are complementary rather than contradictory to each other.

A second (not opposing) explanation is that MTL is involved in both the A/B task and the A/nonA task, but performs a different function in each. Consider that the two training methods likely lead to different category representations (Goldstone, 1996), and the two tasks have different demands. The A/B task leads to a formation of a symmetric representation of two contrasting categories (two prototypes) together with their category label, emphasizing their distinct features (Goldstone, Steyvers, & Rogosky, 2003). During the testing phase, each stimulus needs to be compared to both prototypes and the category that is closer to the current stimulus is selected as a response. The A/nonA task likely leads to a formation of a representation of a single category with its characteristic features (Corneille, Goldstone, Queller, & Potter, 2006). During the testing phase, the stimulus is compared to this single representation. If it is sufficiently similar, it is categorized as a member of the category, if it is not sufficiently similar, it is categorized as a nonmember. Because there are a number of differences in the cognitive demands of the two tasks, let us consider how different functions of the MTL could support these demands.

In order to support episodic memory, MTL needs to perform two complementary functions, pattern separation and pattern completion (Kesner & Hopkins, 2006; O'Reilly & McClelland, 1994; O'Reilly & Rudy, 2000). Pattern separation refers to the ability to extract distinct details of a new episode to avoid catastrophic interference from highly similar (overlapping) previous or future episodes. Pattern completion refers to the ability to recall a complete episode based on a partial cue. One can speculate that the participant may use the ease of pattern completion as a measure of a stimulus category membership in the A/nonA task; the more cues (prototypical features) present in the stimulus, the easier it is reconciled with the complete pattern (the prototype). On the other hand, both pattern separation and pattern completion likely play a role in the A/B task. The pattern separation mechanism seems to be integral for the A/B task through its extraction of discriminative features, creating a representation of the two contrasting categories. Pattern completion may be needed in order to select the more similar prototype and to extract the category label. Although pattern separation and pattern completion mechanisms seem to be based on different subregions of MTL (Bakker, Kirwan, Miller, & Stark, 2008; Kirwan & Stark, 2007), it is not yet clear whether they can account for the results observed here for the A/B task and the A/nonA task. Kirwan, Stark and colleagues found two hippocampal regions - CA3 and dentate gyrus - biased towards pattern separation, and the rest of MTL biased towards pattern completion, an activation pattern that does not map clearly on the A/B versus A/nonA distinction observed here.

Another distinction of memory functions is that of recollection, familiarity, and novelty. Recollection is a retrieval accompanied by specific contextual details, while familiarity and novelty refers to a feeling that an item has been (or not been) previously encountered, without the contextual details of a recall. Considering the demands of the two tasks, one can argue that performance in the A/nonA task can be supported by familiarity or novelty alone, as all prototypical features should be more familiar than all non-prototypical features. On the other hand, performance in the A/B task requires the participant to recall details of learning in order to extract the appropriate category label as all features are equally familiar. Recollection has been shown to critically depend on intact hippocampus and posterior parahippocampus while familiarity-based judgements have been shown to be relatively spared in patients with hippocampal and parahippocampal damage (Holdstock, 2005; Tendolkar et al., 1999; Westerberg et al., 2006). The involvement of anterior hippocampus in the A/nonA task may then reflect novelty detection that has been previously ascribed to this region (Daselaar, Fleck, & Cabeza, 2006; B. A. Strange & Dolan, 2006).

The role of striatum in prototype learning

Not fully expected was the role of striatum in the A/nonA task. The striatum has been implicated in nondeclarative category learning in a number of studies (Nomura et al., 2007; Poldrack et al., 2001; Seger & Cincotta, 2002, 2005; D. Shohamy, Myers, Onlaor, & Gluck, 2004), but it has not been reported in the A/nonA task before. The results reported here suggest that the A/nonA task not only recruits striatum to a larger extent than the A/B task, but also that it can support successful categorization as early as after 20 training trials. This is a novel finding with respect to the A/nonA task and one that needs further investigation. Striatum has been implicated in gradually learning stimulus-response-outcome relationships (Knowlton, Mangels, & Squire, 1996; Packard & Knowlton, 2002; Seger & Cincotta, 2005) and processing feedback (Cincotta & Seger, 2007; Little, Shin, Sisco, & Thulborn, 2006; Maddox & Ing, 2005; Seger & Cincotta, 2005). Neither of these functions was of use in the A/nonA task, yet striatal activation was greater during the A/nonA task than A/B task and was predictive of a correct response.

One mechanism by which the striatum may support A/nonA learning is by its role in extracting regularities across multiple experiences in probabilistic learning (Poldrack, Prabhakaran, & Gabrieli, 1999; D. Shohamy, Myers, Kalanithi, & Gluck, 2008; Wilkinson & Jahanshahi, 2007). Because prototypical features were presented more frequently than non-prototypical features during training, these features could be extracted and expected during testing. Rodriguez and colleagues (Rodriguez, Aron, & Poldrack, 2006) found that the activation of ventral striatum increased parametrically with prediction error. Such a signal could be used for successful discrimination of categorical (small prediction error) from noncategorical (large prediction error) stimuli. In support of this claim, we found two loci in the right ventral striatum that showed greater activation during noncategorical (nonA) than categorical (A) stimulus presentation, although the size of the two loci (21 and 13 voxels) did not reach the statistical threshold. The data presented here suggest that besides its traditional role in stimulus-outcome learning and feedback processing, the striatum may additionally support category learning by extracting expected distribution in a set of stimuli in the absence of outcome or corrective feedback.

The role of the striatum in the A/B task cannot be answered based on the current data. The striatal learning system for the stimulus-outcome association is considered slow and incremental (Ashby, Alfonso-Reese, Turken, & Waldron, 1998), typically dominating performance late in learning, after the initial domination by the hippocampal system (Poldrack et al., 2001; Seger & Cincotta, 2005). Although we did not observe a striatal contribution to the A/B task here, based on the previous literature, it could be recruited eventually, after enough trials have passed.

The role of learning mode in the A/nonA task

It is likely that performance in the A/nonA task is supported by different learning systems when the category representation is acquired incidentally than intentionally. For example, imaging studies including incidental A/nonA tasks consistently report decreased occipital activation for categorical patterns (typically as the sole activation site) while findings from intentional A/nonA tasks are less consistent with each other, and usually different from the incidental findings (Aizenstein et al., 2000; Reber, Gitelman, Parrish, & Mesulam, 2003; Reber, Stark, & Squire, 1998b). Regarding the role of the medial temporal lobe in the A/nonA task, it is possible that it is recruited (in addition to other learning systems) during intentional, but not incidental learning. To support this claim, MTL involvement has been reported in neuroimaging studies in intentional, but not incidental A/nonA tasks (Aizenstein et al., 2000; Reber, Gitelman, Parrish, & Mesulam, 2003; Reber, Stark, & Squire, 1998b) and the neuropsychological studies reporting relatively spared learning in amnesia used incidental learning conditions (Bozoki, Grossman, & Smith, 2006; Keri, Kalman, Kelemen, Benedek, & Janka, 2001; Knowlton & Squire, 1993). We also found evidence for a role of the striatum in the intentional A/nonA task previously unreported in the incidental A/nonA task. As a commonality between the incidental and the intentional A/nonA task, we found some evidence that perceptual learning, reported consistently in the incidental version of the A/nonA task, may also contribute to learning in the intentional A/nonA task.

Generalization

It is important to note that the results presented here provide a snapshot of application of category knowledge early in learning and further research is needed to determine the generalizability beyond this scope. First, we did not address how the contrasting A/B or isolated A category representations are initially generated. Comparing training in the two tasks directly is complicated by differences in the presentation methods and stimuli used, so imaging only the testing phase of the A/nonA task has been the standard so far (Aizenstein et al., 2000; Reber, Stark, & Squire, 1998a; Reber, Wong,

& Buxton, 2002). To help us better understand the initial representation formation, a manipulation involving only differences in the category label and/or instruction during training would need to be developed, perhaps in line with research of Goldstone and colleagues (Corneille, Goldstone, Queller, & Potter, 2006; Goldstone, Steyvers, & Rogosky, 2003). Second, we studied prototype learning at a relatively early stage. The relative contribution of different learning systems and their supporting neural structures likely changes during the course of learning. For instance, as noted above, several category learning studies that used feedback training (like our A/B task) have shown that during the course of learning, participants first rely on the medial temporal lobe, but slowly shift towards the striatal system (Poldrack et al., 2001; Poldrack & Packard, 2003; Seger & Cincotta, 2006). It is possible that the A/B prototype task would exhibit such a shift after extensive training as well. Third, the category structures and stimuli used here differ from the typical A/nonA task, and we used different contrasts than reported previously. Previous research on the A/nonA task identified task-related activations by contrasting activation for the categorical (A) stimuli and the non-categorical (non-A) stimuli. We also tested the contrast of categorical (A) stimuli with non-categorical (non-A) stimuli during the A/nonA task, but no activation locus exceeded our criterion of statistical significance. One possibility is that there were not enough trials (training or testing) to reveal the distinction. Another possibility that the previously reported contrast of categorical (A) and noncategorical (nonA) items is specific for the dot pattern stimuli and/or is present only when non-categorical items are random, unrelated patterns.[21]

Future directions

A strong test of functional relevance of a neural region for a cognitive task is parametric modulation of the region's activity with parametric changes in the task. In the current context, regions supporting the A/B and the A/nonA prototype learning tasks should show a modulation of their activity with the distance of a stimulus to the category

[21] It is also important to note that the two studies that included intentional A/nonA task (Aizenstein et al., 2000; Reber, Gitelman, Parrish, & Mesulam, 2003), both using dot-pattern stimuli, differed in the reported activation sites (there was in fact no overlap).

prototype. Importantly, two different signatures are to be expected for the two tasks. In the A/B task, the representation of the two categories (prototypes) is symmetrical, and the modulation should thus be symmetrical about the category boundary. In the A/nonA task, only one category exists. Regions supporting categorization in this task should thus modulate their activation monotonically with the distance of a stimulus from the category A prototype. The graphical representation of these two predictions is depicted in Figure 6.9. Hints of such a relationship have been reported in the literature. For the A/B task, DeGutis and D'Esposito (2007) have compared easy (far from the boundary) and difficult (near the boundary) stimuli from category A and category B separately. No differences were found between activation to the A stimuli and the B stimuli (suggesting symmetrical representation along the category boundary), but a number of regions were identified that were preferentially active for easy or for difficult stimuli on either side of the bound. In the A/nonA task, differences have been reported between activation to A stimuli compared to non A stimuli (Aizenstein et al., 2000; Reber, Gitelman, Parrish, & Mesulam, 2003), although no finer-grade parametric modulation has been reported in the literature so far. Therefore, it is possible that even if a neural region supported performance in both tasks, it may serve a different function in each. We have attempted to test these predictions by calculating signal change separately for stimuli at four distances from the prototype A (distance 0-2, distance 3-4, distance 6-7 and distance 8-10) for striatal and MTL loci identified in the conjunctive task versus baseline contrast and in the direct A/B versus A/nonA contrast. However, due to low number of trials for each event, these trends were not identifiable. In future research, we would like to address this question.

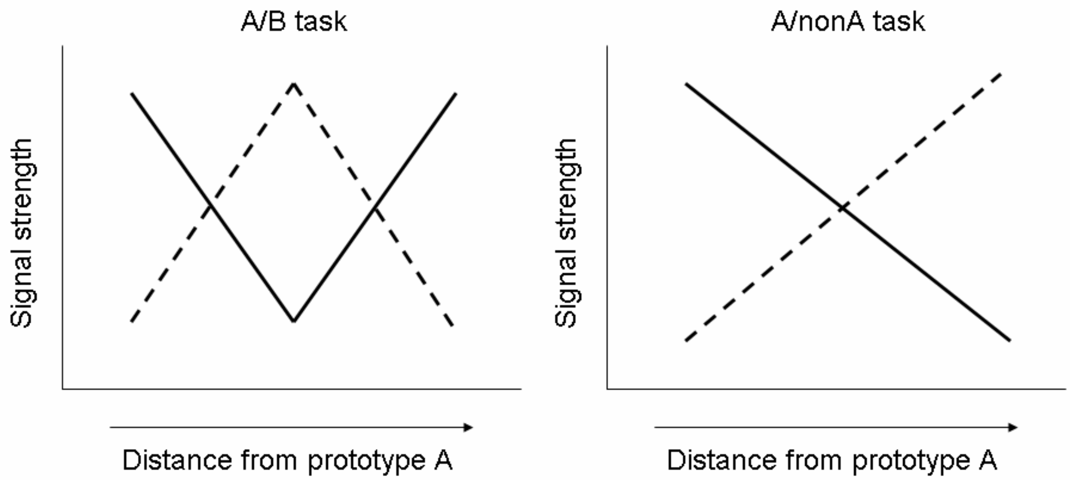

FIGURE 6.9. PREDICTED PARAMETRIC MODULATION OF SIGNAL STRENGTH IN AREAS SUPPORTING THE A/B TASK (LEFT) AND A/NONA TASK (RIGHT). Solid line: Signal decreases as the distance from either prototype (A/B task) or category A prototype (A/nonA task) increases. Dashed line: Signal increases as the distance increases.

Conclusion – the interplay of learning systems

We examined the cognitive and neural processes involved in A/B and A/nonA prototype learning using within subject methodology while simultaneously controlling external variables, such as different category structures, learning modes, and fMRI contrasts that have hindered comparison across existing studies. Based on the current and existing data, we argue existence of dissociable prototype learning pathways. The performance in the A/B task is likely mediated by explicit episodic memory processes based on medial temporal lobe, while performance in the A/nonA task is mediated at least in part by a different system, based (in intentional learning) on striatum and potentially posterior cortices. Based on the previous literature, we speculate that incidental A/nonA learning may differ from both A/B learning and intentional A/nonA learning and is primarily based on changes in processing categorical stimuli within the visual cortex, a form of perceptual learning similar to that observed in perceptual priming (Schacter & Buckner, 1998). The behavioral end-results of the incidental and intentional A/nonA learning seem however similar.

One can speculate that all prototype learning systems likely play an important and complementary role in concept acquisition, as everyday prototype learning experience contains elements of both A/B and A/nonA tasks. We need to learn both characteristic features of a category, as well as features that best differentiate that category from related categories. Each system has its own strengths and limitations. The A/nonA learning can proceed automatically, without intention (incidental learning) and without supervision (Posner & Keele, 1968). Perceptual coherence of the category exemplars seems to be a major limitation in concept learnability (Bozoki, Grossman, & Smith, 2006; Casale & Ashby, in press). Acquiring concepts such as carrot or apple can be supported by the A/nonA type of learning. The A/B prototype learning requires supervision, but allows one to form categories that are less perceptually coherent and make inferences that are not based solely on perceptual similarity. The concept of fruits and vegetables is better suited for the A/B training. While typically operating in parallel, these prototype learning systems are dissociable when demands of the task are tuned to suit one system versus the other (as demonstrated here) or when damage to one system hinders learning of specific task versions (as supported by the neuropsychological literature). Importantly, rather than the category structure itself, the framing and context of the task, such as whether a category is learned in isolation or in contrast to another category, plays a crucial role in recruiting the complementary learning systems.

Chapter 7: General discussion

MULTIPLE CATEGORY LEARNING SYSTEMS

Throughout this dissertation, we discussed the role of four learning and memory systems of the brain in category learning. We found that different systems seem to be dominating in different category learning tasks. A tentative conclusion about the relationships between different category learning tasks and the four memory systems is presented in Table 7.1.

Memory type	Key neural structure	Cognitive functions/tasks	Category learning task	Category learning mechanism
Working memory	Frontal cortex	Maintenance and manipulation of information	Rule-based task	Hypothesis testing
Procedural habit memory	Striatum	Skills and habits; Associative S-R learning	Information-integration task ? Intentional A/nonA task	Association between regions in the perceptual space and category labels
Declarative memory	Medial temporal lobe	Recall and recollection	Prototype A/B task	Exemplar memorization, feature binding
Perceptual representation memory	Sensory cortex	Priming	Incidental A/nonA task ? Intentional A/nonA task	Perceptual memory for categorical items

Table 7.1. Multiple memory systems in category learning tasks.

During the last decade, a number of studies have mapped the cognitive and neural processes involved in rule-based and information-integration category learning in quite some detail (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby, Ell, & Waldron, 2003; Ashby & Maddox, 2005; Filoteo, Maddox, & Davis, 2001b; Filoteo et al., 2005; Maddox & Ashby, 2004; Maddox, Ashby, & Bohil, 2003; Maddox, Ashby, Ing, & Pickering, 2004; Maddox & Ing, 2005; Nomura et al., 2007). In this dissertation, the empirical studies presented in Chapters 2 and 3 (Zeithamova & Maddox, 2006, 2007) helped to establish and understand the role of working memory in rule-based learning. The neural and cognitive processes involved in prototype learning are less understood, and only recently, the dissociation between the A/B prototype task and the A/nonA prototype task started to set attention. The empirical studies presented in Chapters 5 and 6 (Zeithamova, Maddox & Schnyer, in preparation) represent one of the first attempts to contrast the A/B task and the A/nonA task on the behavioral and neural level. Although we could not resolve all antinomies in the existing prototype literature, we were at least partially successful in consolidating some of them. We demonstrated that the A/B task and the A/nonA task rely on dissociable cognitive and neural processes. The A/B task relies primarily on the medial temporal lobe (MTL) dependent declarative memory system, whilte the A/nonA task relies primarily on the sensory cortex based perceptual representation memory and/or the striatum based procedural memory.

Understanding how different learning and memory systems support different categorization tasks have several important theoretical and practical implications. From the theoretical standpoint, by accepting that different categorization tasks may be supported by different learning systems, we were able to sort through the existing literature and interpret previously contradictory findings in the new light. We were also able to propose and conduct new experiments, testing the multiple system theory, and greatly increasing our understanding of cognitive and neural processes underlying categorization. Additionally, once the relationship between different category learning tasks and the memory systems that underlie them is well understood, the categorization tasks themselves can become important tools for studying the memory systems themselves. By realizing that category learning should rely on the same mechanisms as other forms of learning and memory, including the neural substrate that supports them, a bridge can be built between the rich categorization literature and the rich memory literature that have been evolving with insufficient interaction with one another (Knowlton, 1999; Poldrack & Foerde, 2008).

The notion of multiple category learning systems also has implications from a practical standpoint. First, we can use the correspondence between categorization tasks and learning systems for neuropsychological diagnosis. Several studies have already embraced this idea. For example, Filoteo et al (2007) demonstrated that current performance in an information-integration task can predict future cognitive decline in nondemented patients with Parkinson's disease; Keri et al (2002) diagnosed early and late stage of Alzheimer's disease with the A/nonA task, as the early stage affects primarily medial temporal lobe (leaving A/nonA task unaffected), but entire cortex is affected in the late stage (leading to an impaired A/nonA learning). Second, we can improve teaching methods and instruction to utilize all learning and memory systems rather than relying exclusively on explicit memorization. This development has already been applied for some years in second language teaching where explicit learning of language rules got de-emphasized while repetition practice received renewed attention. Using multiple memory systems in a given task can greatly improve speed and quality of performance.

Although we have greatly advanced in our understanding of different category learning systems, we need to ask the same question that has gained recently attention in the memory systems research: Why do we have multiple category learning systems and how do they interact? Although advancing the answer will take a number of years and is beyond the scope of this dissertation, we will discuss a few promising hints in the next section.

CHARACTERISTICS OF THE LEARNING SYSTEMS

Perhaps the easiest answer to the question of why we have multiple category learning systems is because they evolved facing multiple category learning tasks. As we noted in Chapter 1, each system has its characteristic properties – advantages and disadvantages – that make it more or less suitable for certain categorization tasks.

Rule generation, dependent on the frontal cortex, and exemplar memorization, dependent on the MTL, are explicit processes under conscious control. Because of the conscious control, they seem to be the default for humans (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Gluck, Shohamy, & Myers, 2002). Procedural stimulus-response learning and perceptual learning are thought to be automatic, implicit forms of learning that may occur outside awareness (Knowlton & Foerde, 2008; Knowlton, Mangels, & Squire, 1996; Reber & Squire, 1994). The suitability of different learning systems to different tasks and learning context follows from their properties. The explicit, working and declarative memory provide a flexible task representation that can transfer or generalize easily; the implicit, procedural and perceptual memory is closely tied to the original context and lacks the flexibility (Ashby, Ell, & Waldron, 2003; Bayley, Frascino, & Squire, 2005; Casale & Ashby, in press; Gabriele & Packard, 2006; Maddox, Bohil, & Ing, 2004; Maddox, Filoteo, Lauritzen, Connally, & Hejl, 2005; Myers et al., 2003; Reber, Knowlton, & Squire, 1996). Explicit learning also can support fast, even one trial learning, while procedural learning is gradual, requiring extensive practice (Ashby & Ell, 2002; Myers et al., 2003; Squire, 2004). Although these properties seem to favor explicit strategies, the speed and flexibility of explicit knowledge comes with a price. While implicit learning occurs essentially automatically without effort, explicit strategies are dependent on limited resources such as attention, working memory, and conscious awareness (Chapters 2 and 3, Waldron & Ashby, 2001). Additionally, some category structures, such as the information-integration structures, are not easily acquired by an explicit strategy.

Besides a particular categorization task, other factors play a role in determining which learning system would dominate in a categorization task. Factors that have demonstrated effects on the preferential use of different category learning systems include individual differences in the preferred strategy (Gluck, Shohamy, & Myers, 2002), task instruction (E. E. Smith, Patalano, & Jonides, 1998), intention to learn (Aizenstein et al., 2000; Reber, Gitelman, Parrish, & Mesulam, 2003; J. D. Smith & Shapiro, 1989), the length of training (Chang & Gold, 2003; Seger & Cincotta, 2006), behavioral manipulations that disfavor one of the systems (Ashby, Maddox, & Bohil, 2002; Maddox, Ashby, & Bohil, 2003), and neurological limitations involving an impairment of some systems (Bayley, Frascino, & Squire, 2005; Bozoki, Grossman, & Smith, 2006; Reber, Knowlton, & Squire, 1996). In this dissertation, we focused our attention on the categorization task as the factor determining which memory/category learning system dominates learning. Determining how other factors modulate the role of the categorization task in category learning has been only touched here (Chapters 2 and 3) and is a topic of future research.

It is important to note that the correspondence between the memory systems, the category learning mechanisms, and the categorization tasks proposed in Table 7.1 should not be taken as absolute. Rather, I propose that a certain category learning mechanism and a corresponding memory system is likely dominant, but not necessarily exclusive, in mediating learning in a given task. If multiple systems are operative in any given task, how do they interact to support behavior? The next section discusses this question.

INTERACTIVE CATEGORY LEARNING SYSTEMS

The striatal procedural system and the medial temporal declarative system

The only two systems whose interaction has been studied both in the general memory systems research and in the context of category learning are the striatal procedural system and the MTL declarative system. In the tradition of memory research, a number of studies showed that these two systems seem to be competitive in nature (for a review, see e.g. Packard & Knowlton, 2002; Poldrack & Packard, 2003). For instance, a rat can solve a maze task based on learning the rewarded place, an MTL dependent type of memory, or based on learning the rewarded turning sequence (or sometimes learning the rewarded cue), a striatal dependent type of memory (Morris, Garrud, Rawlins, & O'Keefe, 1982; Packard & Teather, 1999). The relative dominance of each system can be affected by a lesion or a pharmaceutical intervention, with lesions to one system sometimes leading to an *improvement* in tasks that rely on the other system (Eichenbaum, Fagan, Mathews, & Cohen, 1988; Morris, Garrud, Rawlins, & O'Keefe, 1982; Packard & McGaugh, 1992, 1996; Packard & Teather, 1997; Schroeder, Wingard, & Packard, 2002). There is also a general strategy shift throughout learning, with place (MTL) responses dominating early in learning and turning (striatal) responses dominating later in learning (Chang & Gold, 2003).

Recently, several neuroimaging studies in humans replicated the interaction of the MTL and the striatum during acquisition of categorization tasks. Specifically, the MTL and striatum seem to be negatively correlated across individuals and across time, with the activity in MTL decreasing and the activity in the striatum increasing with time (Poldrack et al., 2001; Seger & Cincotta, 2006). Both memory and categorization research thus suggest that the striatal procedural system and the MTL declarative system interact throughout learning in a competitive manner, with the MTL dominating early in learning and the striatum dominating later in learning. In this dissertation, we studied one task, the A/B task, primarily dependent on the declarative memory. Based on the literature reviewed above, it is possible that the A/B task may also become dependent on the procedural, striatal system at later stages of learning. As we only included 20 training trials, we were not able to test for the transition from the MTL to the striatum based learning in our data, but future research may address this question.

The striatal procedural system and the frontal hypothesis testing system

Another two systems whose interaction has been proposed in categorization literature is the interaction between the hypothesis testing system and the procedural learning systems proposed in COVIS (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). COVIS assumes that both systems attempt and learn each categorization task encountered. On each trial, both systems generate a response, with the strength (confidence) of the response based on the previous history of success. The systems' responses then compete to determine the final output (the observable response), with the system producing the strongest response winning out. The explicit, frontal hypothesis testing system initially dominates, but when no verbalizable rule providing good performance is found, the striatal procedural system starts dominating over time as the gradual trial-by-trial learning increases its accuracy.

Although further research is needed, I did not find evidence for bilaterally competitive nature of these two systems. First, consider the findings from the memory literature about competition between the striatal and the MTL system. One signature of the competitive nature of the interaction between the two systems was that impairment to one system facilitated learning mediated by the other system. Analogous evidence for the competition between the hypothesis testing system and striatal procedural system has been found only unidirectionally. When behavioral manipulations adversely affect the procedural system, the hypothesis testing system remains dominative throughout learning even for information-integration category structures (Ashby, Maddox, & Bohil, 2002; Maddox, Ashby, & Bohil, 2003). However, we found no evidence that the same would be true for manipulations affecting the hypothesis testing system. In Chapters 2 and 3, we would expect that in the working memory conditions, participants should rely preferentially on the unaffected procedural system. Instead, the participants continued to rely on the unsuccessful hypothesis-testing system. It seems the bias towards the explicit hypothesis-testing strategies is hard to overcome. In the future, perhaps a combination of an information-integration task with an additional instruction manipulation (such as used in e.g. Kemler-Nelson, 1984; E. E. Smith, Patalano, & Jonides, 1998; J. D. Smith & Shapiro, 1989) may be able to promote non-explicit strategies in such circumstances.

Second finding questioning the competitive nature of the two systems comes from computational modeling of the COVIS theory. Consider trial-by-trial resolution of the competition between the two systems. COVIS proposes that the final output (observable response) is generated from the responses of the two subsystems using a winner-takes-all method. An alternative is a "cooperative" nature of this resolution. On the level of the computational model, this can be achieved by determining the final output as a weighted sum of the responses of the two subsystems. Across several applications of the computational model COVIS, we found that the implementation using the weighted sum ("cooperative") resolution provided superior model fit compared to the winner-takes-all ("competitive") resolution (Maddox, Filoteo, & Zeithamova, under review; Zeithamova, Filoteo, Simmons, Maddox, & Paulus, 2007). In summary, based on the data currently available, the frontal hypothesis-testing system and the striatal procedural system seem cooperative rather than competitive in nature.

And the questions for future research

For both interactions discussed above – the declarative memory with the procedural memory and the hypothesis-testing system with the procedural memory, time seems of essence, with a shift from explicit strategies to procedural learning with length of training. Although in this dissertation we only focused on immediate performance evaluation, it is important to note that the retention period also plays an important role. As we discussed in Chapter 1, individual exemplars play an increasingly smaller role and the abstracted prototypes or rules an increasing larger role when category knowledge is tested with a delay after acquisition (Homa, 1973; Homa & Little, 1985; Posner & Keele, 1970; Reed, 1972). Although the procedural and neural mechanism of this shift is not yet well described, it likely involves changes within and/or between supporting memory systems. Similarly, working memory cannot support retention of a categorization rule from session to session; another long-term memory mechanism needs to be involved if what was learned about categories should last. Currently, our understanding of how category learning and representation changes across larger time scales is very incomplete and further research is needed in this area.

Very little is also known about the interactions of the perceptual learning system and any other system. We can speculate that perceptual memory contributes to category learning, providing support to other category learning mechanisms, whenever perceptual similarity within a category is greater than perceptual similarity between categories. The first computational model of perceptual learning in categorization was proposed by Keri and colleagues (2002). We may hope that future category learning models will be able to incorporate this perceptual learning model as a subunit to more complex models.

Finally, we can ask whether the list of memory systems in category learning is now complete. Probably, it is not. Besides the four memory systems that were previously implicated in category learning and thus considered in this dissertation, other memory systems exist (see e.g. Squire, 2004, for a recent review). Two neural structures that have a demonstrated role in learning and memory, but were not discussed here, are cerebellum and amygdala. Cerebellum has been associated with certain types of classical conditioning (McCormick & Thompson, 1984) and motor learning (Ito, 2000). Although there are currently no theories of category learning that would involve the cerebellum, we may find ourselves revising this view in a few years as recent findings suggest that the cerebellum may play a previously underappreciated role in cognition (Paquier & Marien, 2005; Tamminga & Vogel, 2005; Thach, 1998). Similarly, the amygdala has been traditionally implicated in emotional learning, such as fear conditioning (Bechara, Tranel, Damasio, & Adolphs, 1995; Berntson, Bechara, Damasio, Tranel, & Cacioppo, 2007). Recent studies, however, also demonstrated the role of the amygdala in modulation of the MTL and striatum contribution to learning and performance (McIntyre, Marriott, & Gold, 2003a, 2003b; Packard & Wingard, 2004). Understanding the resolution of the competition between the MTL and the striatal learning system based on the task context may thus not be possible unless we include the amygdala into the equation.

Even within the four learning systems that we considered, each system may contribute to category learning via several distinct mechanisms. In this dissertation, we saw two hints of these dissociations within a system. First, in Chapter 6, we found activation in the medial temporal lobe in both the A/B task and the A/nonA task. The specific locus of this activation, however, differed between the tasks, probably reflecting the difference in the specific learning mechanisms active in each task. Second, the striatal system has been traditionally implicated in slow trial-and-error learning, where feedback plays an important role (Ashby, Queller, & Berretty, 1999; Seger & Cincotta, 2005; Daphna Shohamy et al., 2004). However, in Chapter 6, we identified striatum in the A/nonA task, implying that the striatum can extract regularities in the stimuli without external feedback. This finding is not entirely surprising as novel stimuli have been reported to induce similar dopamine release in the striatum as a reward (Lind et al., 2005; Pierce, Crawford, Nonneman, & Mattingly, 1990; Williams, Rolls, Leonard, & Stern, 1993), and activation of the striatum in response to novely was also found in humans using neuroimaging (Berns, Cohen, & Mintun, 1997). It is thus likely that striatum can gradually learn regularities and build expectation across a series of stimuli, even in the absence of external feedback. Such mechanism can support category learning but is distinct from the feedback mediated stimulus-response mapping implied in the information-integration learning. Additionally, anterior striatum is reciprocially connected with the prefrontal cortex as a part of the attentional network. It thus likely plays yet another role in category learning, this time in supporting rule-based learning by mediating attention switches between different stimulus dimensions (Ashby, Noble, Filoteo, Waldron, & Ell, 2003; Brown & Marsden, 1988). Teasing apart distinct mechanisms of category learning with striatum and within MTL is the next step in categorization research.

References

Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., Nebes, R. D., Larson, J. K., Ursu, S., et al. (2000). Complementary category learning systems identified using event-related functional MRI. *Journal of Cognitive Neuroscience, 12*(6), 977-987.

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120*(1), 3-19.

Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin & Review, 8*(4), 629-647.

Anderson, J. R., & Fincham, J. M. (1996). Categorization and sensitivity to correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(2), 259-277.

Ashby, F. G. (1992). Multidimensional models of categorization. In *Multidimensional models of perception and cognition.* (pp. 449-483). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

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

Ashby, F. G., & Casale, M. B. (2003). The cognitive neuroscience of implicit category learning. In L. Jimenez (Ed.), *Attention and implicit learning.* (pp. 109-141). Amsterdam, Netherlands: John Benjamins Publishing Company.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences, 5*(5), 204-210.

Ashby, F. G., & Ell, S. W. (2002). Single versus multiple systems of category learning: Reply to Nosofsky and Kruschke (2001). *Psychonomic Bulletin & Review, 9*, 175-180.

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

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

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

Ashby, F. G., & Maddox, W. T. (1998). *Stimulus categorization* (Vol. 3). New York: Academic Press.

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology, 56*, 149-178.

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., Noble, S., Filoteo, J. V., Waldron, E. M., & Ell, S. W. (2003). Category learning deficits in Parkinson's disease. *Neuropsychology, 17*(1), 115-124.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences, 9*(2), 83-89.

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61*, 1178-1199.

Ashby, F. G., & Spiering, B. J. (2004). The Neurobiology of Category Learning. *Behavioral & Cognitive Neuroscience Reviews, 3*(2), 101-113.

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.

Ashby, F. G., & Waldron, E. M. (2000). The neuropsychological bases of category learning. *Current Directions in Psychological Science, 9*(1), 10-14.

Baddeley, A. D. (1995). Working memory. In M. S. Gazzaniga (Ed.), *Cognitive neurosciences.* (pp. 755-764): The MIT Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control.* (pp. 28-61): Cambridge University Press.

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern Separation in the Human Hippocampal CA3 and Dentate Gyrus. *Science, 319*(5870), 1640-1642.

Barsalou, L. W. (1987). *The instability of graded structure: implications for the nature of concepts*. Cambridge: Cambridge University Press.

Bayley, P. J., Frascino, J. C., & Squire, L. R. (2005). Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature, 436*(7050), 550-553.

Bechara, A., Tranel, D., Damasio, H., & Adolphs, R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science, 269*(5227), 1115-1118.

Bench, C. J., Frith, C. D., Grasby, P. M., & Friston, K. J. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia, 31*(9), 907-922.

Berns, G. S., Cohen, J. D., & Mintun, M. A. (1997). Brain regions responsive to novelty in the absence of awareness. *Science, 276*(5316), 1272-1275.

Berntson, G. G., Bechara, A., Damasio, H., Tranel, D., & Cacioppo, J. T. (2007). Amygdala contribution to selective dimensions of emotion. *Social Cognitive and Affective Neuroscience, 2*(2), 123-129.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 36*(2), 209-231.
Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology, 79*(2), 251-272.
Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 121-133.
Bower, G., & Trabasso, T. (1963). Reversals prior to solution in concept identification. *Journal of Experimental Psychology, 66*(4), 409-418.
Bozoki, A., Grossman, M., & Smith, E. E. (2006). Can patients with Alzheimer's disease learn a category implicitly? *Neuropsychologia, 44*(5), 816-827.
Brainard, D. H. (1997). Psychophysics softward for use with MATLAB. *Spatial Vision, 10*, 433-436.
Brooks, L. (1978). *Nonanalytic concept formation and memory for instances*. Hillsdale, NJ: Erlbaum.
Brown, R. G., & Marsden, C. D. (1988). Internal versus external cues and the control of attention in Parkinson's disease. *Brain, 111*, 23-45.
Bruner, J. S., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New York: Wiley.
Campbell, F. W., & Kulikowski, J. J. (1966). Orientational selectivity of the human visual system. *Journal of Physiology, 187*(2), 437-445.
Casale, M. B., & Ashby, F. G. (in press). A role for the perceptual representation memory system in category learning.
Chang, Q., & Gold, P. E. (2003). Switching memory systems during learning: Changes in patterns of brain acetylcholine release in the hippocampus and striatum in rats. *Journal of Neuroscience, 23*(7), 3001-3005.
Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition, 30*(3), 353-362.
Cincotta, C. M., & Seger, C. A. (2007). Dissociation between striatal regions while learning to categorize via feedback and via observation. *Journal of Cognitive Neuroscience, 19*(2), 249-265.
Cocchini, G., Logie, R. H., Sala, S. D., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition, 30*(7), 1086-1095.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: L. Erlbaum Associates.
Conway, A. R. A., & Engle, R. W. (1996). Individual Differences in Working Memory Capacity: More Evidence for a General Capacity Theory. *Memory, 4*(6), 577-590.
Cooper, L. A., Schacter, D. L., Ballesteros, S., & Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(1), 43-57.

Corneille, O., Goldstone, R. L., Queller, S., & Potter, T. (2006). Asymmetries in categorization, perceptual discrimination, and visual search for reference and nonreference exemplars. *Memory & Cognition, 34*(3), 556-567.

Curran, T. (1995). On the neural mechanisms of sequence learning. *Psyche: An Interdisciplinary Journal of Research on Consciousness, 2*(12).

D'Esposito, M., & Postle, B. R. (1999). The dependence of span and delayed-response performance on prefrontal cortex. *Neuropsychologia, 37*(11), 1303-1315.

D'Esposito, M., Postle, B. R., Stuss, D. T., & Knight, R. T. (2002). The organization of working memory function in lateral prefrontal cortex: Evidence from event-related functional MRI. In *Principles of frontal lobe function.* (pp. 168-187). New York, NY, US: Oxford University Press.

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping, 8*(2-3), 109-114.

Daselaar, S. M., Fleck, M. S., & Cabeza, R. (2006). Triple Dissociation in the Medial Temporal Lobes: Recollection, Familiarity, and Novelty. *J Neurophysiol, 96*(4), 1902-1911.

DeGutis, J., & D'Esposito, M. (2007). Distinct mechanisms in visual category learning. *Cognitive, Affective & Behavioral Neuroscience, 7*(3), 251-259.

Dosher, B. A., & Lu, Z.-L. (1999). Mechanisms of perceptual learning. *Vision Research, 39*(19), 3197-3221.

Dougherty, M. R. P., & Hunter, J. E. (2003). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica, 113*(3), 263-282.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to bootstrap.* New York.: Chapman & Hall.

Eichenbaum, H., Fagan, A., Mathews, P., & Cohen, N. J. (1988). Hippocampal system dysfunction and odor discrimination learning in rats: Impairment of facilitation depending on representational demands. *Behavioral Neuroscience, 102*(3), 331-339.

Eldridge, L. L., Masterman, D., & Knowlton, B. J. (2002). Intact implicit habit learning in Alzheimer's disease. *Behavioral Neuroscience, 116*(4), 722-726.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 127,* 107-140.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature, 407*(6804), 630-632.

Filoteo, J. V., Maddox, W. T., & Davis, J. D. (2001a). A possible role of the striatum in linear and nonlinear category learning: Evidence from patients with Hungtington's disease. *Behavioral Neuroscience, 115*(4), 786-798.

Filoteo, J. V., Maddox, W. T., & Davis, J. D. (2001b). Quantitative modeling of category learning in amnesic patients. *Journal of the International Neuropsychological Society, 7*(1), 1-19.

Filoteo, J. V., Maddox, W. T., Salmon, D. P., & Song, D. D. (2005). Information-Integration Category Learning in Patients With Striatal Dysfunction. *Neuropsychology, 19*(2), 212-222.

Filoteo, J. V., Maddox, W. T., Salmon, D. P., & Song, D. D. (2007). Implicit Category Learning Performance Predicts Rate of Cognitive Decline in Nondemented Patients With Parkinson's Disease. *Neuropsychology, 21*(2), 183-192.

Filoteo, J. V., Maddox, W. T., Simmons, A. N., Ing, A. D., Cagigas, X. E., Matthews, S., et al. (2005). Cortical and subcortical brain regions involved in rule-based category learning. *Neuroreport: For Rapid Communication of Neuroscience Research, 16*(2), 111-115.

Flannagan, M. J., Fried, L. S., & Holyoak, K. J. (1986). Distributional expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12*(2), 241-256.

Folstein, J. R., & Van Petten, C. (2004). Multidimensional Rule, Unidimensional Rule, and Similarity Strategies in Categorization: Event-Related Brain Potential Correlates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(5), 1026-1044.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A Comparison of Primate Prefrontal and Inferior Temporal Cortices during Visual Categorization. *Journal of Neuroscience, 23*(12), 5235-5246.

Furmanski, C. S., & Engel, S. A. (2000). An oblique effect in human primary visual cortex. *Nature Neuroscience, 3*(6), 535-536.

Gabriele, A., & Packard, M. G. (2006). Evidence of a role for multiple memory systems in behavioral extinction. *Neurobiology of Learning and Memory, 85*(3), 289-299.

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the 'weather prediction' task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory, 9*(6), 408-418.

Goldman-Rakic, P. S. (1987). *Circuitory of the prefrontal cortex and the regulation of behavior by representational knowledge*. Bethesda, MD: American Physiological Society.

Goldman-Rakic, P. S. (1990). Cortical localization of working memory. In J. L. McGaugh & N. M. Weinberger (Eds.), *Brain organization and memory: Cells, systems, and circuits.* (pp. 285-298): Oxford University Press.

Goldman-Rakic, P. S. (1998). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. In A. C. Roberts & T. W. Robbins (Eds.), *Prefrontal cortex: Executive and cognitive functions.* (pp. 87-102): Oxford University Press.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition, 24*(5), 608-628.

Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition, 31*(2), 169-180.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Heeley, D. W., & Timney, B. (1988). Meridional anisotropies of orientation discrimination for sine wave gratings. *Vision Research, 28*(2), 337-344.
Hikosaka, O., Sakamoto, M., & Usui, S. (1989). Functional properties of monkey caudate neurons: III. Activities related to expectation of target and reward. *Journal of Neurophysiology, 61*(4), 814-832.
Holdstock, J. S. (2005). The role of the human medial temporal lobe in object recognition and object discrimination. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology, 58*(3), 326-339.
Homa, D. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology, 101*(1), 116-122.
Homa, D., & Little, J. (1985). The abstraction and long-term retention of ill-defined categories by children. *Bulletin of the Psychonomic Society, 23*(4), 325-328.
Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7*(6), 418-439.
Hull, C. L. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs, 28*(1), 1-86.
Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98*(3), 352-376.
Ito, M. (2000). Mechanisms of motor learning in the cerebellum. *Brain Research, 886*(1), 237-245.
Janowsky, J. S., Shimamura, A. P., Kritchevsky, M., & Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral Neuroscience, 103*(3), 548-560.
Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage, 17*(2), 825-841.
Jonides, J., Reuter-Lorenz, P. A., Smith, E. E., Awh, E., Barnes, L. L., Drain, M., et al. (1996). Verbal and spatial working memory in humans. In D. L. Medin (Ed.), *Psychology of learning and motivation: Advances in research and theory, Vol. 35.* (pp. 43-88): Academic Press.
Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*(1), 47-70.
Kemler-Nelson, D. G. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior, 23*(6), 734-759.
Keri, S. (2003). The cognitive neuroscience of category learning. *Brain Research Reviews, 43*(1), 85-109.
Keri, S., Janka, Z., Benedek, G., Aszalas, P., Szatmary, B., Szirtes, G., et al. (2002). Categories, prototypes and memory systems in Alzheimer's disease. *Trends in Cognitive Sciences, 6*(3), 132-136.

Keri, S., Kalman, J., Kelemen, O., Benedek, G., & Janka, Z. (2001). Are Alzheimer's disease patients able to learn visual prototype? *Neuropsychologia, 39*(11), 1218-1223.
Kesner, R. P., & Hopkins, R. O. (2006). Mnemonic functions of the hippocampus: A comparison between animals and humans. *Biological Psychology, 73*(1), 3-18.
Kirwan, C. B., & Stark, C. E. L. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learn. Mem., 14*(9), 625-633.
Knowlton, B. J. (1999). What can neuropsychology tell us about category learning? *Trends in Cognitive Sciences*, *3*(4), 123-124.
Knowlton, B. J., & Foerde, K. (2008). Neural representations of nondeclarative memories. *Current Directions in Psychological Science, 17*(2), 107-111.
Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273*(5282), 1399-1402.
Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science, 262*(5140), 1747-1749.
Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory, 1*(2), 106-120.
Knowlton, B. J., Squire, L. R., Paulsen, J. S., Swerdlow, N. R., & Swenson, M. (1996). Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology, 10*(4), 538-548.
Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.
Labov, W. (1973). The boundaries of words and their meanings. In C. J. Bailey & R. Shuy (Eds.), *New ways of analysing variations in English*. Washington, D.C.: Georgetown university press.
Lakoff, G. (1986). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.
Laurienti, P. J. (2004). Deactivations, Global Signal, and the Default Mode of Brain Function. *Journal of Cognitive Neuroscience, 16*(9), 1481-1483.
Lawrence, A. D., Sahakian, B. J., Hodges, J. R., & Rosser, A. E. (1996). Executive and mnemonic functions in early Huntington's disease. *Brain: A Journal of Neurology*, *119*(5), 1633-1645.
Lawrence, A. D., Watkins, L. H. A., Sahakian, B. J., Hodges, J. R., & Robbins, T. W. (2000). Visual object and visuospatial cognition in Huntington's disease: Implications for information processing in corticostriatal circuits. *Brain: A Journal of Neurology*, *123*(7), 1349-1364.
Lei, M., & Zhansheng, C. (2003). Rule-based categorization strategy and example-based categorization strategy in categorization. *Acta Psychologica Sinica, 35*(1), 29-36.
Leng, N. R., & Parkin, A. J. (1988). Double dissociation of frontal dysfunction in organic amnesia. *British Journal of Clinical Psychology*, *27*(4), 359-362.
Levy, A. S., & Heshka, S. (1973). Similarity and the false recognition of prototypes. *Bulletin of the Psychonomic Society, 1*(3), 181-183.

Levy, R., Friedman, H. R., Davachi, L., & Goldman-Rakic, P. S. (1997). Differential activation of the caudate nucleus in primates performing spatial and nonspatial working memory tasks. *Journal of Neuroscience, 17*(10), 3870-3882.

Lind, N. M., Gjedde, A., Moustgaard, A., Olsen, A. K., Jensen, S. B., Jakobsen, S., et al. (2005). Behavioral response to novelty correlates with dopamine receptor availability in striatum of Gottingen minipigs. *Behavioural Brain Research, 164*(2), 172-177.

Little, D. M., Shin, S. S., Sisco, S. M., & Thulborn, K. R. (2006). Event-related fMRI of category learning: Differences in classification and feedback networks. *Brain and Cognition, 60*(3), 244-252.

Little, D. M., & Thulborn, K. R. (2005). Correlations of cortical activation and behavior during the application of newly learned categories. *Cognitive Brain Research, 25*(1), 33-47.

Little, D. M., & Thulborn, K. R. (2006). Prototype-distortion category learning: A two-phase learning process across a distributed network. *Brain and Cognition, 60*(3), 233-243.

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica, 75*(1), 55-74.

Love, B. C. (2003). The multifaceted nature of unsupervised category learning. *Psychonomic Bulletin & Review*, *10*(1), 190-197.

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

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163-203.

Maddox, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Perception & Psychophysics, 61*(2), 354-375.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics, 53*, 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, and Cognition*, *29*(4), 650-662.

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., Bohil, C. J., & Ing, A. D. (2004). Evidence for a procedural-learning-based system in perceptual category learning. *Psychonomic Bulletin & Review, 11*(5), 945-952.

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.
Maddox, W. T., Filoteo, J. V., Lauritzen, J. S., Connally, E., & Hejl, K. D. (2005). Discontinuous Categories Affect Information-Integration but Not Rule-Based Category Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(4), 654-669.
Maddox, W. T., Filoteo, J. V., & Zeithamova, D. (under review). Computational models inform clinical science and assessment: An application to category learning in striatal-damaged patients.
Maddox, W. T., & Ing, A. D. (2005). Delayed Feedback Disrupts the Procedural-Learning System but Not the Hypothesis-Testing System in Perceptual Category Learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31*(1), 100-107.
Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering Minds: The Default Network and Stimulus-Independent Thought. *Science, 315*(5810), 393-395.
McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6*(4), 462-472.
McCloskey, M. E., & Glucksberg, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology, 11*(1), 1-37.
McCormick, D. A., & Thompson, R. F. (1984). Cerebellum: Essential involvement in the classically conditioned eyelid response. *Science, 223*(4633), 296-299.
McIntyre, C. K., Marriott, L. K., & Gold, P. E. (2003a). Cooperation between memory systems: Acetylcholine release in the amygdala correlates positively with performance on a hippocampus-dependent task. *Behavioral Neuroscience, 117*(2), 320-326.
McIntyre, C. K., Marriott, L. K., & Gold, P. E. (2003b). Patterns of brain acetylcholine release predict individuals differences in preferred learning strategies in rats. *Neurobiology of Learning and Memory, 79*(2), 177-183.
Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207-238.
Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7*(5), 355-368.
Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7*(4), 241-253.

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society, 7*(3), 283-284.
Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 775-799.
Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(2), 275-292.
Mishkin, M., Malamut, B., & Bachevalier, J. (1984). *Memories and habits: Two neural systems*. New York: Guilford Press.
Morris, R. G., Garrud, P., Rawlins, J. N., & O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature, 297*(5868), 681-683.
Moss, M., Mahut, H., & Zola-Morgan, S. (1981). Concurrent discrimination learning of monkeys after hippocampal, entorhinal, or fornix lesions. *J. Neurosci., 1*(3), 227-240.
Murphy, G. L., Ahn, W.-k., Goldstone, R. L., Love, B. C., Markman, A. B., & Wolff, P. (2005). The Study of Concepts Inside and Outside the Laboratory: Medin Versus Medin. In *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin.* (pp. 179-195). Washington, DC, US: American Psychological Association.
Myers, C. E., Shohamy, D., Gluck, M. A., Grossman, S., Kluger, A., Ferris, S., et al. (2003). Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *Journal of Cognitive Neuroscience, 15*(2), 185-193.
Narayanan, N. S., Prabhakaran, V., Bunge, S. A., Christoff, K., Fine, E. M., & Gabrieli, J. D. E. (2005). The Role of the Prefrontal Cortex in the Maintenance of Verbal Working Memory: An Event-Related fMRI Analysis. *Neuropsychology, 19*(2), 223-232.
Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*(1), 1-32.
Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., et al. (2007). Neural Correlates of Rule-Based and Information-Integration Visual Category Learning. *Cerebral Cortex, 17*(1), 37-43.
Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39-57.
Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(1), 87-108.
Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 700-708.
Nosofsky, R. M. (1992a). *Exemplar-based approach to relating categorization, identification, and recognition*. Hillsdale, NJ: Erlbaum.

Nosofsky, R. M. (1992b). Exemplars, prototypes, and similarity rules. In A. F. Healy, S. M. Kosslyn & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes.* (pp. 149-167). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 282-304.
Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of 'multiple-system' phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7*(3), 375-402.
Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review, 9*, 169-174.
Nosofsky, R. M., Kruschke, J. K., & McKinley, S. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 211-233.
Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin & Review, 5*(3), 345-369.
Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). A rule-plus-exception model of classification learning. *Psychological Review, 101*, 53-79.
Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 247-255.
O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus, 4*(6), 661-682.
O'Reilly, R. C., & Rudy, J. W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus, 10*(4), 389-397.
Olsson, H., Wennerholm, P., & Lyxzen, U. (2004). Exemplars, Prototypes, and the Flexibility of Classification Models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 936-941.
Orban, G. A., Vandenbussche, E., & Vogels, R. (1984). Human orientation discrimination tested with long stimuli. *Vision Research, 24*(2), 121-128.
Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience, 25,* 563-593.
Packard, M. G., & McGaugh, J. L. (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: Further evidence for multiple memory systems. *Behavioral Neuroscience, 106*(3), 439-446.
Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory, 65*(1), 65-72.
Packard, M. G., & Teather, L. A. (1997). Double dissociation of hippocampal and dorsal-striatal memory systems by posttraining intracerebral injections of 2-amino-5-phosphonopentanoic acid. *Behavioral Neuroscience, 111*(3), 543-551.

Packard, M. G., & Teather, L. A. (1999). Dissociation of multiple memory systems by posttraining intracerebral injections of glutamate. *Psychobiology, 27*(1), 40-50.

Packard, M. G., & Wingard, J. C. (2004). Amygdala and 'emotional' modulation of the relative use of multiple memory systems. *Neurobiology of Learning and Memory, 82*(3), 243-252.

Paquier, P. F., & Marien, P. (2005). A synthesis of the role of the cerebellum in cognition. *Aphasiology, 19*(1), 3-19.

Patalano, A. L., Smith, E. E., Jonides, J., & Koeppe, R. A. (2001). PET evidence for multiple strategies of categorization. *Cognitive, Affective & Behavioral Neuroscience, 1*(4), 360-370.

Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Pierce, R. C., Crawford, C. A., Nonneman, A. J., & Mattingly, B. A. (1990). Effect of forebrain dopamine depletion on novelty-induced place preference behavior in rats. *Pharmacology, Biochemistry and Behavior, 36*(2), 321-325.

Poldrack, R. A. (2002). Neural systems for perceptual skill learning. *Behavioral and Cognitive Neuroscience Reviews, 1*(1), 76-83.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature, 414*(6863), 546-550.

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience & Biobehavioral Reviews, 32*(2), 197-205.

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia, 41*(3), 245-251.

Poldrack, R. A., Prabhakaran, S. C. A., & Gabrieli, J. D. E. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13*, 564-574.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*(3), 353-363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*(2), 304-308.

Postle, B. R., Jonides, J., Smith, E. E., & Corkin, S. (1997). Spatial, but not object, delayed response is impaired in early Parkinson's disease. *Neuropsychology, 11*(2), 171-179.

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral and Brain Sciences, 28*(1), 1-49.

Raghavachari, S., Kahana, M. J., Rizzuto, D. S., Caplan, J. B., Kirschen, M. P., Bourgeois, B., et al. (2001). Gating of Human Theta Oscillations by a Working Memory Task. *J. Neurosci., 21*(9), 3175-3183.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proc Natl Acad Sci U S A, 98*(2), 676-682.

Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience, 15*(4), 574-583.
Reber, P. J., Knowlton, B. J., & Squire, L. R. (1996). Dissociable properties of memory systems: Differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral Neuroscience, 110*(5), 861-871.
Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning & Memory, 1*(4), 217-229.
Reber, P. J., & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. *Behavioral Neuroscience, 113*(2), 235-242.
Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998a). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory, 5*(6), 420-428.
Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998b). Cortical areas supporting category learning identified using functional MRI. *Proceedings of National Academy of Sciences, 95*(2), 747-750.
Reber, P. J., Wong, E. C., & Buxton, R. B. (2002). Comparing the brain areas supporting nondeclarative categorization and recognition memory. *Cognitive Brain Research, 14*(2), 245-257.
Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*(3), 382-407.
Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning.* (pp. 21-59). New York, NY, US: Cambridge University Press.
Rips, L. J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General, 122*(4), 468-486.
Robinson, A. L., Heaton, R. K., Lehman, R. A. W., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal of Consulting & Clinical Psychology, 48*, 605-614.
Rodriguez, P. F., Aron, A. R., & Poldrack, R. A. (2006). Ventral-striatal/nucleus-accumbens sensitivity to prediction errors during classification learning. *Human Brain Mapping, 27*(4), 306-313.
Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4*(3), 328-350.
Rosch, E. (1975a). Cognitive reference points. *Cognitive Psychology, 7*, 532-547.
Rosch, E. (1975b). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192-233.
Rosch, E. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382-439.
Rosch, E. (1978). *Principles of categorization*. Hillsdale: Erlbaum.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573-605.

Rosch, E., & Moore, T. E. (1973). On the internal structure of perceptual and semantic categories. In *Cognitive development and the acquisition of language.* Oxford, England: Academic Press.

Roth, E. M., & Mervis, C. B. (1983). Fuzzy set theory and class inclusion relations in semantic categories. *Journal of Verbal Learning & Verbal Behavior, 22*(5), 509-525.

Schacter, D. L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. *Annals of the New York Academy of Sciences, 608*, 543-571.

Schacter, D. L. (1994). *Priming and multiple memory systems: Perceptual mechanisms of implicit memory*. Cambridge, Massachusetts: MIT Press.

Schacter, D. L., & Buckner, R. L. (1998). On the relations among priming, conscious recollection, and intentional retrieval: Evidence from neuroimaging research. *Neurobiology of Learning and Memory, 70*(1), 284-303.

Schacter, D. L., Cooper, L. A., & Delaney, S. M. (1990). Implicit memory for visual objects and the structural description system. *Bulletin of the Psychonomic Society, 28*(4), 367-372.

Schacter, D. L., Cooper, L. A., Tharan, M., & Rubens, A. B. (1991). Preserved Priming of Novel Objects in Patients with Memory Disorders. *Journal of Cognitive Neuroscience, 3*(2), 117-130.

Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Current Opinion in Neurobiology, 17*(2), 171-176.

Schroeder, J. P., Wingard, J. C., & Packard, M. G. (2002). Post-training reversible inactivation of hippocampus reveals interference between memory systems. *Hippocampus, 12*(2), 280-284.

Schultz, W., Apicella, P., Romo, R., Scarnati, E., Houk, J. C., Davis, J. L., et al. (1995). Context-dependent activity in primate striatum reflecting past and future behavioral events. In *Models of information processing in the basal ganglia.* (pp. 11-27): The MIT Press.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 164-464.

Seger, C. A., & Cincotta, C. M. (2002). Striatal activity in concept learning. *Cognitive, affective, & behavioral neuroscience, 2*(2), 149-161.

Seger, C. A., & Cincotta, C. M. (2005). The Roles of the Caudate Nucleus in Human Classification Learning. *Journal of Neuroscince, 25*(11), 2941-2951.

Seger, C. A., & Cincotta, C. M. (2006). Dynamics of Frontal, Striatal, and Hippocampal Systems during Rule Learning. *Cerebral Cortex, 16*(11), 1546-1555.

Seger, C. A., Poldrack, R. A., Prabhakaran, V., Zhao, M., Glover, G. H., & Gabrieli, J. D. E. (2000). Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. *Neuropsychologia, 38*(9), 1316-1324.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4-27.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*(13).

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of "dot-pattern" classification and recognition. *Journal of Experimental Psychology: General, 121*, 278-304.

Shohamy, D., Myers, C. E., Grossman, S., Sage, J., Gluck, M. A., & Poldrack, R. A. (2004). Cortico-striatal contributions to feedback-based learning: Converging data from neuroimaging and neuropsychology. *Brain: A Journal of Neurology, 127*(4), 851-859.

Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience & Biobehavioral Reviews, 32*(2), 219-236.

Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004). Role of the Basal Ganglia in Category Learning: How Do Patients With Parkinson's Disease Learn? *Behavioral Neuroscience, 118*(4), 676-686.

Slotnick, S. D., & Schacter, D. L. (2006). The nature of memory related activity in early visual areas. *Neuropsychologia, 44*(14), 2874-2886.

Smith, E. E. (1995). Concepts and categorization. In E. E. Smith & D. N. Osherson (Eds.), *Thinking: An invitation to cognitive science, Vol. 3 (2nd ed.).* (pp. 3-33). Cambridge, MA, US: The MIT Press.

Smith, E. E., & Grossman, M. (2008). Multiple systems of category learning. *Neuroscience & Biobehavioral Reviews, 32*(2), 249-264.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6*(1), 11-20.

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65*(2), 167-196.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1411-1436.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 3-27.

Smith, J. D., & Minda, J. P. (2001). Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(4), 984-1002.

Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 800-811.

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.

Smith, J. D., & Shapiro, J. H. (1989). The occurence of holistic categorization. *Journal of Memory & Language, 28*(4), 386-399.

Smith, S. (2002). Fast robust automated brain extraction. *Human Brain Mapping, 17*(3), 143-155.
Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory, 82*(3), 171-177.
Stanton, R. D., Nosofsky, R. M., & Zaki, S. R. (2002). Comparisons between exemplar similarity and mixed prototype models using a linearly separable category structure. *Memory & Cognition, 30*(6), 934-944.
Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652-654.
Stewart, N., & Chater, N. (2002). The effect of category variability in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(5), 893-907.
Strange, B. A., & Dolan, R. J. (2006). Anterior medial temporal lobe in human cognition: Memory for fear and the unexpected. *Cognitive Neuropsychiatry, 11*(3), 198-218.
Strange, W., Keeney, T., Kessel, F. S., & Jenkins, J. J. (1970). Abstraction over time of prototypes from distortions of random dot patterns: A replication. *Journal of Experimental Psychology, 83*(3), 508-510.
Tamminga, C. A., & Vogel, M. (2005). Images in Neuroscience: The cerebellum. *American Journal of Psychiatry, 162*(7), 1253-1253.
Tendolkar, I., Schoenfeld, A., Golz, G., Fernández, G., Kühl, K.-P., Ferszt, R., et al. (1999). Neural correlates of recognition memory with and without recollection in patients with Alzheimer's disease and healthy controls. *Neuroscience Letters, 263*(1), 45-48.
Thach, W. T. (1998). What is the role of the cerebellum in motor learning and cognition? *Trends in Cognitive Sciences, 2*(9), 331-337.
Trabasso, T., & Bower, G. (1964). Presolution reversal and dimensional shifts in concept identification. *Journal of Experimental Psychology, 67*(4), 398-399.
Tracy, J. I., Mohamed, F., Faro, S., Pinus, A., Tiver, R., Harvan, J., et al. (2003). Differential brain responses when applying criterion attribute versus family resemblance rule learning. *Brain and Cognition, 51*(3), 276-286.
Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247*(4940), 301-306.
Tversky, A., & Hutchinson, J. W. (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review, 93*(1), 3-22.
Varela, J. A., Sen, K., Gibson, J., Fost, J., Abbott, L. F., & Nelson, S. B. (1997). A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *J Neurosci, 17*(20), 7926-7940.
Verbeemen, T., Vanpaemel, W., Pattyn, S., Storms, G., & Verguts, T. (2007). Beyond exemplars and prototypes as memory representations of natural concepts: A clustering approach. *Journal of Memory and Language, 56*(4), 537-554.
Vygotsky, L. S. (1962). *Thought and language*. Oxford, England: Wiley.
Wagner, A. D., Gabrieli, J. D. E., & Verfaellie, M. (1997). Dissociations between familiarity processes in explicit recognition and implicit perceptual memory.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 23(2), 305-323.
Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review, 8*(1), 168-176.
Westerberg, C. E., Paller, K. A., Weintraub, S., Mesulam, M. M., Holdstock, J. S., Mayes, A. R., et al. (2006). When Memory Does Not Fail: Familiarity-Based Recognition in Mild Cognitive Impairment and Alzheimer's Disease. *Neuropsychology*, *20*(2), 193-205.
Westwater, H., McDowall, J., Siegert, R., Mossman, S., & Abernethy, D. (1998). Implicit learning in Parkinson's disease: Evidence from a verbal version of the serial reaction time task. *Journal of Clinical and Experimental Neuropsychology, 20*(3), 413-418.
Wickens, J. (1993). *A theory of the striatum*. New York: Pergamon Press.
Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology, 8*(2), 227-233.
Wilkinson, L., & Jahanshahi, M. (2007). The striatum and probabilistic implicit sequence learning. *Brain Research, 1137*, 117-130.
Williams, G. V., Rolls, E. T., Leonard, C. M., & Stern, C. (1993). Neuronal responses in the ventral striatum of the behaving macaque. *Behavioural Brain Research, 55*(2), 243-252.
Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1047-1060.
Wilson, C. J. (1995). *The contribution of cortical neurons to the firing pattern of striatal spiny neurons*. Cambridge: MIT Press.
Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003). Single Neurons in the Monkey Hippocampus and Learning of New Associations. *Science, 300*(5625), 1578-1581.
Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.
Worsley, K. J. (2001). Statistical analysis of activation images. In P. Jezzard, P. M. Matthews & S. Smith (Eds.), *Functional MRI: An Introduction to Methods.*
Zaki, S. R., & Nosofsky, R. M. (2004). False prototype enhancement effects in dot pattern categorization. *Memory & Cognition*, *32*(3), 390-398.
Zaki, S. R., & Nosofsky, R. M. (2007). A high-distortion enhancement effect in the prototype-learning paradigm: Dramatic effects of category learning during test. *Memory & Cognition, 35*(8), 2088-2096.
Zaki, S. R., Nosofsky, R. M., Jessup, N. M., & Unversagt, F. W. (2003). Categorization and recognition performance of a memory-impaired group: Evidence for single-system models. *Journal of the International Neuropsychological Society, 9*(3), 394-406.
Zeithamova, D., Filoteo, J. V., Simmons, A. N., Maddox, W. T., & Paulus, M. P. (2007). *Category learning systems: Combining behavior, computational model and fMRI.*

Paper presented at the Annual Meeting of the Cognitive Neuroscience Society, New York.
Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition, 34*(2), 387-398.
Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition, 35*(6), 1380-1398.
Zeithamova, D., & Maddox, W. T. (under review). Learning mode and exemplar sequencing in unsupervised category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

VITA

Dagmar Zeithamova entered Charles University in Prague in 1996, with major in psychology and minor in logic and economics. She received the degree of Master of Art from Charles University in Prague in February, 2003. In September 2003, she entered the Graduate School at The University of Texas at Austin.

Permanent address: Manesova 37, Praha 2, 120 00, Czech republic

This dissertation was typed by the author.