

THE NEURAL CORRELATES OF EXPLICIT CATEGORIZATION

By

Marci A. Flanery

Dissertation

Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

May, 2005

Nashville, Tennessee

Approved:

Professor Thomas J. Palmeri

Professor Randolph Blake

Professor Timothy P. McNamara

Professor Rene Marois

Professor David C. Noelle

Copyright © 2005 by Marci A. Flanery
All Rights Reserved

To Fozzi Flanery, an old soul and beloved companion.

ACKNOWLEDGEMENTS

This dissertation would not have been possible were it not for Patty Roberts – who always had good advice for me, Pat Burns – who always had faith in me, and Fozzi Flanery and Bartok Roberts – who were always happy to see me. But, I would be remiss to limit my acknowledgements to such a short list. So many people helped me in so many ways over the years that it is not possible to describe their individual contributions here. These people include my fellow graduate students: Moria Smoski, Kim Curby, Ken Sobel, Christian Luhmann, Julia High, Jessecae Marsh, Emmy Grossman, Jon Holbrook; my fellow lab members: Naren Robey, Tim Vickery, Talia Valdez, and Eva Greco; colleagues and collaborators: Jim Stefansic, Patrick Henry, George Holburn, Danielle Nikolaicyzk, Vicky Morgan, Sean Jewett, Jody Leavell, woo-Kyoung Ahn, Craig Stark, Amy Shelton, Jo-Anne Bachorowski, and Andrew Tomarken; as well as family: Phyllis & Henry Keil, Robert Roberts, Rod Roberts; and friends: Larry A. Belger and Dr. James Jackson. I would also like to offer long overdue thanks to several individuals from my hometown of Halstead, Kansas: Clarence Scharff, Dr. Jack and Mrs. Jane Welch, and Larry and Caroline Williams. The functional imaging costs related to these studies as well as a portion of my financial support while at Vanderbilt were provided by the following grants: *fMRI Studies of Perceptual Categorization* from Vanderbilt University Discovery Grants Program; *Functional Brain Imaging of Cognitive Processes: Studies of "Perceptual" and "Probabilistic" Category Learning* from Vanderbilt GCRC Project #1054 – NIH GCRC Grant M01 RR00095; *Rules and Instances in Perceptual Categorization*, R01 MH61370, from the National Institute of Mental Health; and *Perceptual Categorization and Memory*, BCS-9910756, from the National Science Foundation. And finally, I offer my thanks and appreciation to the members of my committee: Randolph Blake, Tim McNamara, Rene Marois, David Noelle, and Tom Palmeri.

TABLE OF CONTENTS

	Page
DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF FIGURES	viii
LIST OF TABLES	ix
Chapter	
I. INTRODUCTION TO AND OVERVIEW OF EXPERIMENTS IN THIS DISSERTATION	1
II. THEORIES OF CATEGORIZATION AND NEURAL SYSTEMS OF LEARNING AND MEMORY.....	5
Exemplar Theories of Categorization and the Medial Temporal Lobe System.....	5
Prototype Theories of Categorization and the Cortico-Cortical System	8
Rule-Based and Hybrid Theories of Categorization and the Fronto-Striatal System.....	10
III. NEURAL CORRELATES ASSOCIATED WITH DIFFERENCES IN CATEGORIZATION TASKS	18
Experiment 1	18
Method	23
Subjects	23
Apparatus and Image Acquisition Parameters	23
Stimuli	23
Prototype-distortion Condition	24
Prototype-distortion Control Condition	24
Probabilistic-cue Condition	24
Probabilistic-cue Control Condition	25
Fixation	25
Procedure	26
Procedure – Early Epoch Condition	26
Runs	26
Blocks	26
Trials	26
Procedure – Late Epoch Condition	27
Late Epoch Pre-training	27
Runs	27
Blocks & Trials	27
Analysis.....	27

Behavioral Data Analysis	27
Functional Imaging Analysis	27
Individual Subjects.....	27
Group Analysis	28
Results.....	29
Behavioral.....	29
Neuroimaging	30
Discussion	30
Experiment 2.....	35
Method	39
Subjects.....	39
Apparatus and Image Acquisition Parameters.....	39
Stimuli	39
Probabilistic Category Structure.....	39
Deterministic Category Structure	40
Control Conditions.....	40
Fixation	40
Procedure	40
Functional Imaging Sessions	40
Functional Imaging Runs.....	41
Trials	41
Analysis	42
Behavioral Data Analysis	42
Functional Imaging Analysis.....	42
Individual Subjects.....	42
Group Analysis	43
Results & Discussion	44
Behavioral.....	44
Neuroimaging	44
Categorization Regions-of-Interest.....	44

IV. CHANGES IN NEURAL CORRELATES OVER THE COURSE OF CATEGORY ACQUISITION.....50

Experiment 3.....	50
Method	55
Subjects.....	55
Apparatus and Image Acquisition Parameters.....	55
Stimuli.....	55
Procedure	56
Pre-Scan Familiarization Session	56
Functional Imaging	56
LEARN Trials.....	57
TEST Trials.....	58
CNTRL Trials	58
Analysis.....	58

Individual Subject Data.....	58
Group Analysis	59
ROI Analyses	59
Maximum Intensity ROI Map.....	59
Rule-Based ROI Map.....	59
MTL ROI Map.....	60
Results & Discussion	60
Behavioral	60
Maximum Intensity ROIs	61
<i>A Priori</i> Rule and MTL ROIs	63
Anatomical Atlas Based Basal Ganglia ROIs.....	68
Attention versus Rules	68
V. GENERAL DISCUSSION	69
Learning Subsystems	72
Basal Ganglia.....	72
Medial Temporal Lobe	73
Sensory Neocortex	73
Mediation by Orbitofrontal, Lateral Prefrontal, and Anterior Cingulate Cortex.....	74
Conclusion	75
Appendices	
A. STIMULI – EXPERIMENT 2.....	77
B. MATERIALS – EXPERIMENT 3.....	78
REFERENCES	83

LIST OF FIGURES

Figure	Page
1. Schematic of COVIS.....	15
2. Block Design for Experiment 1	25
3. Mean percent correct in the four categorization conditions of Experiment 1.....	29
4. EPOCH x TASK Interaction – right inferior parietal lobule	33
5. Performance in the four test conditions of Experiment 2	44
6. LTxFX Interaction – left thalamus	46
7. Schematic of block design used in Experiment 3	57
8. Mean percent correct for categorization performance for each RUN (1 through 8) of the task	61
9. Activation of Left Hippocampal Complex	67
10. Neural areas active with categorization relative to control tasks common across Experiments 1, 2, and 3.....	71
11. Card Stimuli for Experiment 2.....	77
12. Rocket stimuli used in Experiment 3	79
13. Lamp stimuli used in Experiment 3	80
14. Sailboat stimuli used in Experiment 3	81
15. Necktie stimuli used in Experiment 3.....	82

LIST OF TABLES

Table	Page
1. Primary connections of the corticostriatohalamacortical loops.....	12
2. Probabilistic-cue combinations for card stimuli	25
3. Experiment 1 Regions of Interest and ANOVA Results	31
4. Experimental Conditions of Experiment 2	38
5. Structure of learning, testing and control trials in Experiment 2.....	42
6. Experiment 2 Regions of Interest and ANOVA Results	45
7. Category Structures used for Experiment 3	56
8. Experiment 3 Regions of Interest and ANOVA Results	62
9. Rule ROIs with Significant Effects.....	64
10. Medial Temporal Activations – Main Effect of RUN	66
11. Neural Areas Active with Categorization relative to Control: Experiments 1, 2, and 3....	70
12. Rocket Dimensions	79
13. Lamp Dimensions	80
14. Sailboat Dimensions	81
15. Necktie Dimensions	82

CHAPTER I

INTRODUCTION AND OVERVIEW OF EXPERIMENTS IN THIS DISSERTATION

Categorization is a fundamental cognitive process. When we judge an object to be a member of a category, we are then able to make inferences about that object based only on its category membership. For instance, when presented with a flower we might immediately classify it as a rose. Based only on this category assignment, we are careful when handling the stem because roses have thorns. Notwithstanding the thorns, we may still incline our head toward the petals because roses smell good. We know that roses have both thorns and a pleasing fragrance based on our prior exposure to many different members of the rose category and it is not necessary to rediscover these common features with each new rose that we encounter. This ability to generalize from accumulated past experience to our present experience allows us to fluidly interact with a cluttered and constantly changing world, guiding our actions to both avoid harm and seek pleasure.

Understanding categorization has been the goal of an extensive amount of research in cognitive psychology. Using behavioral (Homa, 1978; Homa & Vosburgh, 1976; Medin & Schaffer, 1978; Posner, Goldsmith, & Welton, 1967; Posner & Keele, 1968; S. K. Reed, 1972), mathematical modeling (Anderson, 1991; Ashby & Gott, 1988; Ashby & Maddox, 1993; Estes, 1986; Hintzman, 1986; Koriat, Goldsmith, & Pansky, 2000; Lamberts, 2000; Nosofsky, 1984; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994; Palmeri & Nosofsky, 2001), and neuropsychological (Alvarez, Zola-Morgan, & Squire, 1995; Knowlton, Mangels, & Squire, 1996; Knowlton & Squire, 1993; Knowlton, Squire, & Gluck, 1994; J. M. Reed, Squire, Patalano, Smith, & Jonides, 1999) techniques, scientists have documented numerous behavioral phenomena associated with categorization and formulated theories to account for these phenomena. However, even with this extensive corpus of knowledge, current research based on behavioral and modeling methods does not support any single theory of categorization above the rest since each of the competing theories is able to account for many of the same categorization results.

One example of this failure to reach consensus can be found with respect to the nature of the neural systems underlying categorization. Although early theories of categorization presumed

a single-categorization system, recent evidence from the fields of behavioral neuropsychology and mathematical modeling suggest that there may be different neural systems associated with different aspects of categorization tasks. However, there are limitations on the conclusions that can be drawn from each these experimental approaches. For instance, behavioral neuropsychological evidence obtained from lesion-deficit models cannot be obtained for all brain regions and the effects of particular lesions are not always consistent across subjects (Price & Friston, 2002). In addition, when degenerate systems exist, there may be no evidence of a deficit associated with a lesion that affects only one of several systems that are sufficient to complete the task (Edelman & Gally, 2001; Price & Friston, 2002). Similarly, evidence from mathematical modeling of behavioral results from empirical studies is not conclusive in that equivalent models proposing different cognitive processes may be constructed and good model fits can often be obtained for ostensibly poor models of cognitive process (Olsson, Wennerholm, & Lyxzen, 2004; Pitt & Myung, 2002). Therefore, although we have learned a great deal about categorization using these research methods, no single research method appears to be sufficient to address the nature of the neural systems supporting categorization.

Although cognitive psychologists have traditionally been interested in understanding psychological process independent of brain structure (Nyberg, 1999), with advances in our knowledge of the relationship between brain structure and psychological function there is now a move to accept that the brain can inform our understanding of the mind. Just as statistically significant differences in behavioral measures such as reaction time and percent correct are presumed to reflect psychologically relevant differences in processing, statistically significant differences in neural activation are also presumed to reflect psychologically relevant differences in processing. In addition, the spatial locations – neural loci – of these differences can also reveal information about psychological process based on our cumulative knowledge of tasks that are associated with differential activation in these same neural regions. Thus, experimentation utilizing functional neuroimaging methods can provide additional, converging, evidence with which to understand psychological process.

In this dissertation, I utilize functional magnetic resonance imaging (fMRI) to address questions about the neural basis of the cognitive processes supporting the explicit learning and testing of perceptual category knowledge. Although the goal of the dissertation is to identify the neural correlates of explicit category learning and testing, the experimental paradigms I have

chosen to explore this issue have all been previously associated with the identification of multiple- or hybrid- categorization systems in the behavioral, behavioral neuropsychological, or modeling literature. The choice of these paradigms is based on the idea that the behavior elicited by these particular tasks can be thought of as representing extremes along a continuum and for that reason, these paradigms should also provide the greatest opportunity to find differential activation either within or between neural regions with respect to differences in experimental conditions. This approach should therefore provide us with insight with respect to the neural processes supporting a range of categorization tasks rather than just those neural processes associated with a single categorization task relative to a non-categorization control.

In Experiments 1 and 2, fMRI is used to identify neural correlates associated with different aspects of categorization tasks such as the stage of learning (early versus late), type of categorization task (prototype-distortion versus probabilistic-cue), category assignment (deterministic versus probabilistic), and type of training (feedback versus exposure). In Experiment 1, I examine variations of two categorization tasks that have been used in behavioral neuropsychological research, the results of which have supported the existence of multiple-categorization systems. Similarities and differences in neural activation associated with performance during both the early and late stages of these two tasks is then evaluated in light of the extant neuropsychological, modeling and fMRI literature. In Experiment 2, I partially replicate the early condition of Experiment 1 controlling for differences in the stimuli. Review of the extant fMRI literature suggests that differences in neural activation associated with different categorization paradigms may actually be associated with the type of learning paradigm (feedback versus exposure) rather than the type of categorization task (prototype-distortion versus probabilistic-cue). Therefore, a second goal of Experiment 2 is to examine whether differences in the type of categorization task (prototype-distortion versus probabilistic-cue), differences in learning paradigm (feedback versus exposure), and differences in category phase (learning versus testing) are associated with differences in functional activation.

In contrast to Experiments 1 and 2, in which the primary focus is on differences in neural activation associated with differences in the categorization tasks, Experiment 3 investigates in greater detail the neural activation associated with performance of a single categorization task over the course of eight learning epochs completed in a single training session. Evidence from mathematical modeling of behavioral data has demonstrated that there is a qualitative shift in

categorization behavior as subjects perform category learning and transfer tasks (Johansen & Palmeri, 2002; Nosofsky, Palmeri, & McKinley, 1994; D. J. Smith & Minda, 1998). Various explanations proffered for this shift have included a change in representation from rules to exemplars (Johansen & Palmeri, 2002), a change in selective attention to features (Nosofsky & Johansen, 2000), and a change in representation from prototypes to exemplars (D. J. Smith & Minda, 1998). In Experiment 3, I examine the pattern of functional activation over the course of category acquisition within neural regions previously associated with rule-based categorization, hypothesis testing, and guessing (Elliott & Dolan, 1998; Elliott, Rees, & Dolan, 1999; Patalano, Smith, Jonides, & Koeppe, 2001) and LTM encoding and retrieval (Squire, Stark, & Clark, 2004) utilizing *a priori* regions-of-interest to examine changes in neural activation consistent with a change in representation from rules-to-exemplars.

Experiments 1, 2, and 3 investigate the neural substrates associated with paradigms from the extant behavioral, fMRI, and modeling literature and discussion with respect to these experiments is initially limited to the neural substrates specifically associated with the experimental manipulations. However, in the final chapter I will return to a question that spans these three experiments that is, what are the neural processes associated with the performance of explicit categorization tasks. The picture that is beginning to emerge from the extant functional imaging literature and the results of the current experiments is one of a dynamic process in which the neural regions associated with categorization include a group of stable underlying regions that are active relative to control conditions across a variety of categorization tasks as well as neural regions that are specifically associated with various components of the categorization task, such as the learning paradigm and stage of acquisition.

This dissertation is presented in five chapters including this introduction. Chapter II reviews the relationship between various models of categorization and the neural systems of learning and memory that may instantiate this cognitive process. Experiments 1 and 2, examining neural differences associated with differences in categorization tasks, are then presented jointly in Chapter III and Experiment 3, examining neural differences associated with a single categorization task over the course of category learning, is presented in Chapter IV. Chapter V concludes with an analysis of activation across the three experiments and a general discussion of what the three experiments may tell us about the neural correlates of explicit category learning and testing.

CHAPTER II
THEORIES OF CATEGORIZATION AND NEURAL SYSTEMS
OF LEARNING AND MEMORY

How is perceptual categorization instantiated in the brain? Do the neural processes supporting categorization change with various aspects of the categorization task? with increased experience? or, do these neural processes change with both task and experience? These questions arise from a body of categorization research that has evolved from two research traditions – empirical studies of normal human behavior and behavioral neuropsychological studies of subjects with brain damage. The goal of the current chapter is to examine how various theories of categorization relate to the neural systems of learning and memory. To this end, three possible views of how categorization may be instantiated in the brain are discussed. In the first view, categorization is dependent on explicit memory processes that are supported by the same medial temporal learning and memory system as recognition and recall. In the second view, categorization is dependent on a nondeclarative memory process that is supported by a cortico-cortical learning and memory system. And, in the third view, categorization is dependent on competition between nondeclarative memory processes and verbalizable rule-based processes that are supported by a fronto-striatal learning and memory system.

Exemplar Theories of Categorization and the Medial Temporal Lobe System

The medial temporal lobe (MTL) structures specifically associated with normal LTM function are often collectively referred to as the hippocampal complex (HC) and can generally be defined as including the parahippocampal cortex, perirhinal cortex, entorhinal cortex, hippocampus (which includes the CA1, CA2, and CA3 subfields), dentate gyrus, fimbria, fornix, and subicular complex (Amaral, 1999). The array of intact and impaired abilities traditionally associated with damage to these anatomical regions within the MTL system is well known and the syndrome is referred to as anterograde amnesia (AA) (Milner, 1968; Milner, Corkin, & Teuber, 1968; Parkin, 1997; Scoville, 1968). AA is characterized by intact short-term memory in conjunction with impaired long-term memory (LTM). Thus, if you introduce yourself and enter

into a brief conversation with a subject with AA (amnesiac), the amnesiac can address you correctly during the conversation since they will be able to retain your name in short-term memory however, 24-hours later you can repeat the exact same conversation since the amnesiac is unable to utilize LTM to recall your name or the conversation from the prior day's encounter. In fact, poor performance on tests of *recall* (the ability to produce a previously seen item when cued by a previously associated item), and *recognition* (the ability to correctly identify a previously seen item as "old" when presented a second time), is a hallmark of the amnesiac syndrome. Recall and recognition tasks test episodic, or declarative, memory which can be defined as conscious, LTM for specific events (Squire et al., 2004) such as where you last placed your car keys or that the University of Kansas Jayhawks won the NCAA Final Four basketball championship in 1988.

The MTL neural system is generally associated with various exemplar models of categorization (Hintzman, 1986; Nosofsky, 1988a, 1991; Nosofsky & Zaki, 1998; Palmeri & Flanery, 1999, 2002). Exemplar models are based on the assumption that each instance of a category is stored as an exemplar in LTM and that these exemplars are accessed in order to make a category decision. Exemplar models generally determine categorization of a target item as a member or nonmember of a category based on similarity. For instance, in the Generalized Context Model (GCM), similarity is based on the similarity of the to-be-categorized item (the target) to the stored exemplars of the category relative to the stored exemplars of all relevant categories (Nosofsky, 1984, 1988b). In general, the more similar a target exemplar is to other stored members of a particular category, the more likely the individual will respond affirmatively that the target item is a member of that category. Nosofsky has demonstrated that the GCM can be used to model both recognition and categorization decisions since, in essence, recognition can be thought of as a case of categorization where the category consists of one item (Nosofsky, 1986, 1988a). Since intact recognition performance depends on an intact MTL (Squire et al., 2004), the relationship between categorization and recognition in the GCM implies that exemplar-based categorization is also dependent on an intact MTL (Nosofsky & Zaki, 1998; Palmeri & Flanery, 1999, 2002).

Despite behavioral and modeling evidence that categorization is dependent on an intact MTL system, evidence of this relationship from functional neuroimaging is relatively sparse. In one study that did identify differential activation of the MTL, Reber and colleagues (Reber,

Gitelman, Parrish, & Mesulam, 2003) used distortions of dot-patterns as stimuli (prototype-distortion paradigm) to compare activation associated with implicit and explicit category learning. In this experiment, one group of subjects learned the category through incidental (implicit) exposure and a second group of subjects learned the category through intentional (explicit) exposure. Subjects in both conditions were then tested on their ability to make categorical judgments with respect to the learned category while functional images were collected (functional images were not collected during the category learning phase of the experiment). In an anatomically defined ROI analysis of the MTL, Reber and colleagues identified an increase in activity in the left anterior hippocampus for the intentional categorization condition relative to the incidental categorization condition.

Aizenstein et al. (Aizenstein et al., 2000) also utilized a prototype-distortion paradigm in a task comparing implicit and explicit exposure learning of categorical stimuli. However, in this study, presenting the different categorical stimuli in different colors and having subjects perform a color-mapping task achieved exposure to the category. In the incidental exposure condition subjects were instructed to attend to the dot patterns as they were presented paying attention only to the color. A pattern would be presented for 1500 ms after which it would change to either red, yellow or blue and remain on the screen for an additional 1500 ms. In both conditions, subjects were instructed to press a key corresponding to the color of the stimulus. In the explicit learning condition subjects were instructed to also look for a pattern within the presented stimuli. In contrast to Reber et al. (2003), subjects in this study were scanned only during the acquisition phase of the task and not during a subsequent transfer test. In this study, Aizenstein and colleagues reported a region of right MTL that showed increased activity relative to baseline for both categorical and noncategorical stimuli in both the incidental and intentional learning conditions.

Finally, Poldrack and colleagues have also reported differential activity in the MTL for a probabilistic-cue categorization task relative to a perceptual-motor baseline task. In this study, the category learning condition was explicit but the feedback was probabilistic rather than deterministic as in the Reber (Reber et al., 2003) and Aizenstein (Aizenstein et al., 2000) studies. In this study, two regions of the left hippocampus were found to have differential activation relative to the baseline task.

Prototype Theories of Categorization and the Cortico-Cortical System

A second learning and memory system that does not depend on an intact MTL is the cortico-cortical (also referred to as neocortical) system (Aggleton & Brown, 1999; Baddeley, Vargha-Khadem, & Mishkin, 2001; McClelland, McNaughton, & O'Reilly, 1995; O'Reilly & Rudy, 2000, 2001). This learning and memory system appears to be important for slow (as opposed to one-trial) learning of items based on multiple exposures to stimuli. The neural substrate underlying this particular system is not as well specified as that underlying the MTL system but would include portions of the occipital, parietal, and cingulate cortex. One candidate substrate for this system would be the second (non-hippocampal) axis of Aggleton and Brown's (1999) dual-axis theory. Aggleton and Brown (1999) have proposed that dual anatomical systems account for the pattern of results reported in amnesiac and animal lesion studies of LTM as measured by tests of recognition and recall. In this system, a hippocampal-anterior thalamic system (part of the medial temporal system previously described) is integral for episodic memory that supports recall and remember-recognition judgments, and a separate perirhinal-medial dorsal thalamic system is capable of supporting know-recognition judgments based on familiarity. The non-hippocampal know-recognition circuit is not dependent on the hippocampus and is dependent on direct projections of the perirhinal cortex to the medial dorsal nucleus of the thalamus. One of the implications of Aggleton and Brown's theory is that if damage is limited to the hippocampal circuit, the subject will experience anterograde amnesia coupled with intact know-recognition.

Whereas AA is associated with impaired explicit memory processes supported by the MTL, it is also associated with spared implicit memory processes that may be supported by the intact cortico-cortical system. Implicit (nondeclarative) memory can be described as unconscious memory for items or events to which the subject has previously been exposed (Squire et al., 2004). Implicit memory is usually assessed via indirect measures such as decreased reaction times for previously encountered items relative to new items. In some types of skill learning tasks, such as tracing a figure by looking at your hand in a mirror, amnesiacs show the same practice effects as normal subjects (evidence of spared implicit memory) even though they will deny having previously completed the task (evidence of impaired explicit memory).

The ability of this cortico-cortical system to support learning and memory in the absence of an intact MTL system has recently been identified in a group of subjects with a syndrome called developmental amnesia (DA). DA arises from ischaemic or anoxic episodes during infancy or early childhood that result in significant neuronal death of the CA1 field of the hippocampus (Baddeley et al., 2001; Vargha-Khadem et al., 1997). Thus, in DA, the hippocampal circuit of Aggleton & Brown's (1999) dual-axis system is disrupted while the cortical learning system supported by the perirhinal-medial dorsal thalamic circuit remains intact. One of the remarkable aspects of DA is that, despite severe impairments in spatial, temporal, and episodic memory, individuals with this disorder are able to acquire semantic knowledge sufficient to progress through elementary and high school (Vargha-Khadem et al., 1997) as well as exhibit some spared recognition memory skills (Baddeley et al., 2001; Broman, Rose, Hotson, & Casey, 1997) thus demonstrating the ability of the cortico-cortical system to acquire information in the absence of an intact MTL system.

Prototype theories of categorization are often associated with the cortico-cortical learning and memory system¹. A basic assumption of prototype theories is that categorization decisions are based on a comparison of the to-be-categorized item to an abstracted central tendency – prototype – of previously seen category members (Homa, 1978; Homa & Vosburgh, 1976; S. K. Reed, 1972). If an individual is asked to determine whether two target stimuli are members of the category, the reaction time and accuracy of the response will depend on the target's similarity to the prototype (or prototypes in a multiple category model) rather than a similarity metric based on a comparison of the target item to each of the previously seen category members and nonmembers as in exemplar theory.

If categorization is dependent on the cortico-cortical learning and memory system rather than the MTL system, then amnesiacs should be unimpaired at categorization. Behavioral evidence supporting this view comes from a series of prototype-distortion categorization studies that have reported AA subjects can categorize items as well as normal control subjects despite being impaired on old-new recognition judgments for the same stimuli when compared to normal age-matched controls (Filoteo, Maddox, & Davis, 2001; Knowlton & Squire, 1993; J. M. Reed et al., 1999; Squire & Knowlton, 1995).

¹ Although prototype theories are often associated with a cortical learning and memory system, there is nothing in these theories *per se* that would require their dependence on a cortical learning system. Prototype theories could also be instantiated via the MTL or fronto-striatal learning systems.

Functional neuroimaging also provides evidence of a cortico-cortical categorization system. In two studies utilizing an implicit exposure paradigm, Reber, Stark & Squire (1998a, 1998b) have consistently found an area of decreased activation in posterior occipital cortex (BA 18/19) associated with the processing of category members relative to both nonmembers – items generated from the same super-prototype as the member stimuli – (Reber, Stark, & Squire, 1998a) and random dot-patterns (Reber, Stark, & Squire, 1998b). They have interpreted this region of deactivation as reflecting more rapid or less effortful processing of categorical stimuli that can, in turn, support categorization.

Rule-Based and Hybrid Theories of Categorization and the Fronto-Striatal System

Rule-based and hybrid models of categorization are often associated with a third neural system of learning and memory, the fronto-striatal system. For purposes of this thesis, the frontal system includes both frontal and prefrontal cortical regions and the striatal system is broadly construed to include the ventral and dorsal striatum (caudate, putamen, nucleus accumbens), the basal ganglia (globus pallidus, substantia nigra pars reticulata, ventral pallidum) and the thalamus (ventral anterior and medial dorsal nuclei). The striatum is innervated by the entire cerebral cortex, including the limbic system, and projects to the basal ganglia, which, in turn project to the thalamic nuclei.

Frontal cortex is most often associated with short-term and/or working memory. In addition to the striatal connections noted above, frontal cortex is highly interconnected with all of the sensory, motor, and subcortical structures of the brain (Miller & Cohen, 2001). This interconnection with other brain structures places the frontal cortex in a position to be involved in many higher-level brain functions. This possibility is born out by the results of a meta-analysis of 275 imaging studies of various perceptual and cognitive tasks (Cabeza & Nyberg, 2000). In this meta-analysis, frontal lobe activation was associated with almost all of the tasks studied. However, despite the large number of fMRI studies reporting frontal lobe activation, the functional organization of this cortical region and how it supports higher-level cognitive function is still debated. Fletcher and Henson (2001) have reviewed the results of fMRI studies of working memory, episodic encoding, and episodic retrieval and have proposed that three divisions of the lateral frontal lobes are associated with three different types of information

processing. They suggest that the ventrolateral regions are active during updating and maintenance of information, the dorsolateral regions are active during selection, manipulation, and monitoring of information, and the anterior lateral cortex is active during the selection of processes and subgoals (Fletcher & Henson, 2001). Similarly, other researchers have also associated ventral regions with maintenance and dorsal regions with manipulation of information (Owen, Milner, Petrides, & Evans, 1996; Petrides, 1996). However, this characterization of the function of frontal cortex is not the only possibility. An alternative organization of frontal cortex based on evidence from single-cell recordings in primates is that the ventro- and dorsolateral regions are associated with memory and attention whereas the orbital and medial regions are associated with inhibition (Goldman-Rakic, 1987). Yet another view, proposed by O'Reilly and colleagues (O'Reilly, Noelle, Braver, & Cohen, 2002), is that the frontal cortex is organized based on the level of a stimulus representation's abstraction with orbital areas encoding featural information and dorsolateral areas encoding more abstract dimensional information.

Recent evidence also suggests frontal cortex is associated with the use of rules (Bunge, Kahn, Wallis, Miller, & Wagner, 2003; Elliott & Dolan, 1998; Elliott et al., 1999; Miller, 2000; Miller & Cohen, 2001; Miller, Freedman, & Wallis, 2002; O'Reilly et al., 2002; Patalano et al., 2001; E. E. Smith, Patalano, & Jonides, 1998; Wallis, Anderson, & Miller, 2001; Wallis & Miller, 2003). For instance, Wallis and colleagues (Wallis et al., 2001; Wallis & Miller, 2003) have demonstrated single-cell activity in the frontal cortex of monkeys with the application of abstract rules. After training the monkeys on two abstract rules, neurons were recorded from dorsolateral, ventrolateral, and orbital prefrontal cortex while monkeys applied the rules. Neurons selective to each of the rules were found throughout these regions. Bunge, et al. (2003) extended this paradigm to humans also identifying areas of frontal and motor cortex associated with association of abstract rules to uncommon objects.

Similar to the research on the role of the frontal cortex, there is no general agreement on a single-role of the striatal system in learning and memory (Joel, Niv, & Ruppin, 2002). Three prominent views that have emerged as to the role of the basal ganglia system in human information processing are that the basal ganglia are primarily involved in serial processing (Gurney, Prescott, & Redgrave, 2001), action selection (Gurney et al., 2001), and/or reinforcement learning (Joel et al., 2002; Schultz, Trembley, & Hollerman, 2003). A fourth possibility recently suggested by Joel and colleagues (2002) is that the basal ganglia act as a

Table 1
Primary Connections of Corticostriathalamocortical Loops

Cortex	Corticostriathalamocortical Loop			
	Oculomotor	Dorsolateral	Orbitofrontal	Limbic
	FEF SEF DLPFC PPC	DLPFC PPC	LOPC	ACC MOFC
Striatum	Caudate	Caudate	Caudate	Ventral
Direct:	SNr, GPi	GPi, SNr	GPi, Snr	V. Pallidum
Indirect:	GPe, STh	GPe, Sth	GPe, STh	
Thalamus	Vapc VAmc DMpm	Vapc Dmpc	Vamc Dmmc	DMmc

FEF, frontal eye field; SEF, supplementary eye field; DLPFC, dorsolateral prefrontal cortex; PPC, posterior parietal cortex; LOPC, lateral orbital prefrontal cortex; ACC, anterior cingulate cortex; MOFC, medial orbitofrontal cortex; SNr, substantia nigra pars reticulata; GPi, internal segment of globus pallidus; GPe, external segment of globus pallidus; STh, subthalamic nucleus; VApC, ventral anterior nucleus pars parvocellularis; VAmc, ventral anterior nucleus pars magnocellularis; DMpm, dorsomedial nucleus pars multiformis; DMpc, dorsomedial nucleus pars parvocellularis; DMmc, dorsomedial nucleus pars magnocellularis (adapted, Afifi & Bergman, 1998).

dimension reduction mechanism as part of a cortico-pallido-thalamo-cortical loop (Joel et al., 2002). Regardless of the functional role of the striatal system, it is generally agreed that this system mediates cortical processing via five functional loops, the corticostriathalamocortical loops, or circuits (Afifi & Bergman, 1998) which project to motor, oculomotor, dorsolateral prefrontal, lateral orbitofrontal prefrontal, and limbic cortex. The primary connections of the four loops that project to frontal cortex are shown in Table 1.

Function of the striatal system is impaired in patients with Parkinson's Disease (PARK) (Packard & Knowlton, 2002). PARK is caused by disruption of the dopaminergic system (dopamine is a neuromodulator) within the striatum. Although PARK is most often identified with its physical effects – tremors and halted or ballistic movements – it has also been associated with deficits in tasks such as stimulus-response associations, delayed alternation, mirror reading, sequence learning, probabilistic-cue classification, cognitive set shifting, and response selection (Packard & Knowlton, 2002). Individuals with PARK are reported to be impaired at early epochs of categorization tasks in which the category feedback structure is probabilistic (Knowlton et al., 1996) but to have no impairments with respect to perceptual (prototype-distortion) categorization (Reber & Squire, 1999).

Lesions of the frontal lobe are a second source of fronto-striatal dysfunction. These lesions, often the result of stroke or accident, disrupt the fronto-striatal learning system by effectively disconnecting one or more of the corticostriatalthalamocortical loops. In particular, lesions of the dorsolateral prefrontal circuit have been associated with deficits in spatial memory and cognitive disturbances whereas lesions of the lateral orbitofrontal loop have been observed to result in impairments in switching behaviors and obsessive-compulsive behaviors (Packard & Knowlton, 2002).

Rule-based theories of categorization are often associated with the fronto-striatal learning and memory system. Historically, the earliest models of categorization adopted a classical theory of categorization in which categories were represented as a set of necessary and sufficient conditions, or rules (Bruner, Goodnow, & Austin, 1956; E. E. Smith & Medin, 1981). However, the classical theories failed to explain many of the behavioral results obtained in categorization experiments and prototype (Minda & Smith, 2002; S. K. Reed, 1972; D. J. Smith & Minda, 1998), exemplar (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984), and decision bound (Ashby & Gott, 1988; Ashby & Maddox, 1993) theories of categorization emerged as early successors to these rule-based theories. However, rule-based categorization is now reemerging as a component of several hybrid models of categorization (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ericksen & Kruschke, 1998; Johansen & Palmeri, 2002; Nosofsky & Palmeri, 1998; Nosofsky, Palmeri et al., 1994).

Neuroimaging has also provided evidence of fronto-striatal activation with categorization. Patalano and colleagues have utilized PET to examine the differential role of rules and exemplars in learning perceptual categories (Patalano et al., 2001; E. E. Smith et al., 1998). Patalano et al. adapted the methodology and behavioral analyses used by Allen and Brooks (1991) in which test items were constructed such that positive test items would be categorized the same and negative test items would be categorized differently based on the categorization strategy (rule application or exemplar similarity) used by the subject (Allen & Brooks, 1991). Stimuli consisted of cartoon animals with 10 binary-valued features. Subjects learned to categorize items using either a rule-based or exemplar-similarity strategy before being scanned. Subjects received feedback during the learning trials but not during the transfer trials and PET data were collected only during transfer trials. Results were interpreted as revealing two distinct neural patterns. During rule application, prefrontal (BA 6 and BA 46) and posterior

parietal (BA 7) areas were active with respect to a passive viewing baseline. The exemplar similarity condition was associated with occipital (BA 17, 18, 19) and cerebellar regions. Patalano et al. also reported a second condition of this experiment in which subjects were given a new rule for each of the testing runs in order to decrease the possibility that subjects reverted to an exemplar strategy for making their decisions in the rule-based condition. The results of this second study largely replicated the first experiment. The results of the two experiments taken together suggest bilateral activation of frontal regions within BAs 9, 44, 46 and BA 6, parietal regions within BAs 7/19, and occipital regions within BA 18 (inferior to the AC/PC line) are associated with a rule-based categorization strategy and the exemplar similarity condition is associated with activations in occipital cortex.

Whereas Patalano and colleagues (Patalano et al., 2001; E. E. Smith et al., 1998) tested whether there were differences in neural activation associated with categories acquired using either rule-based or exemplar-similarity *strategies*, Ashby and colleagues (Ashby et al., 1998) have proposed a hybrid model of category learning called COVIS (COMPETITION BETWEEN VERBAL AND IMPPLICIT SYSTEMS) that is based on *competition* between two categorization systems. Ashby and colleagues propose that human categorization can be accounted for by two underlying neuropsychological systems – a verbal rule-based categorization system and an implicit categorization system. Although previous researchers may have associated certain neuropsychological processes with categorization, COVIS proposes a complete neuropsychological circuit associated with each of the two competing categorization systems within the model.

Figure 1 presents a schematic of the major components and pathways of COVIS (Ashby et al., 1998). In the implicit system (large broken lines in Figure 1), a representation of the stimulus is computed in extrastriate (inferotemporal) cortex. Extrastriate cortex then projects to the tail of the caudate nucleus where the association between the stimulus and category response is made. The caudate tail either learns a category decision bound through the feedback process or, alternatively, associates a category response with a region of perceptual space (Ashby & Waldron, 1999). After the category response is determined, the response choice is communicated to prefrontal cortex via an activation circuit that includes the globus pallidus and the thalamus.

The verbal system in COVIS (Ashby et al., 1998) (small broken lines in Figure 1) includes the anterior cingulate and prefrontal cortex. The model specifies that the anterior

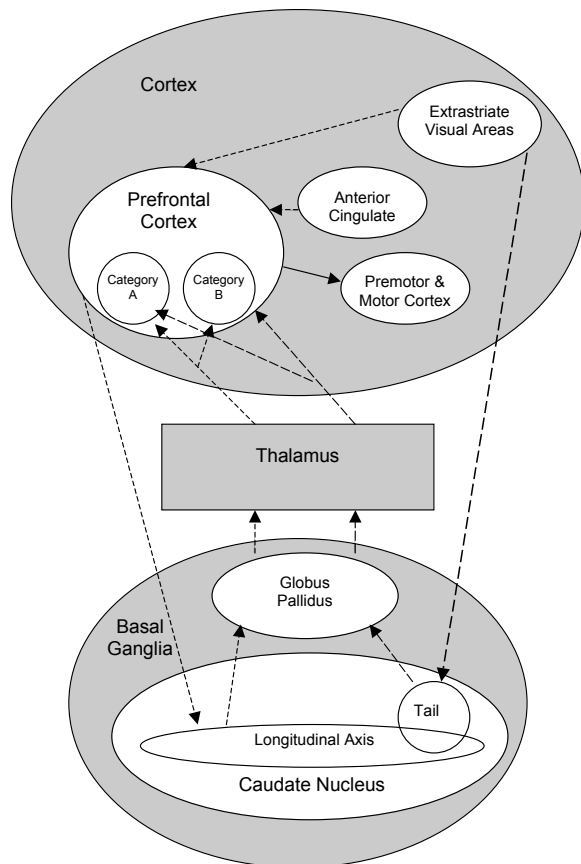


Figure 1. Schematic of COVIS. In COVIS both an implicit system (large broken lines) and a verbal system (small broken lines) act in competition to make category decisions (adapted from Ashby, et al., 1998).

cingulate is responsible for rule selection and that the selected rule is then communicated to the prefrontal cortex. Connections from prefrontal cortex activate the longitudinal axis of the caudate. The role of the caudate in the verbal system is to mediate rule switching and learn the appropriate response criteria. As in the implicit system, once a response category is determined, the response choice is communicated to prefrontal cortex via an activation circuit that includes the globus pallidus and the thalamus.

At this point in the process, both the implicit and the verbal system will have computed a category response but these responses may not be the same, resulting in competition between the two systems (Ashby et al., 1998). In COVIS, the strongest response wins out, but a neural mechanism for resolving this response competition is not specified. One possible neural mechanism to implement this winner-take-all

strategy could be lateral inhibition within the striatum prior to communicating the response on to motor cortex (solid line) since both circuits include this structure.

Seiger and Cincotta (2002) have used functional imaging to test one of the underlying assumptions of the COVIS model. In their study, Seiger and Cincotta examined activation in the striatum during three visual concept learning tasks: a verbal, rule-based, task in which rectangles could be categorized based on their height and width; an implicit task in which the integration of the length and angle of two lines determined the appropriate category response; and a novel-implicit task in which stimuli could be categorized based on the diameter of a circle and angle of the diameter line (Seiger & Cincotta, 2002). The COVIS model (Ashby et al., 1998) would predict activation in the head of the caudate associated with verbal rule application and in the tail of the caudate associated with implicit category learning. Although significant activation of the

striatum was observed in both the verbal and implicit categorization tasks tested by Seger and Cincotta, differences in activation between the head and tail of the caudate relative to the characteristics of the two tasks was not found. However, it should be noted that in this experiment Seger and Cincotta did not collect functional images during the early acquisition phase of category learning. Failure to find a difference in caudate activation between the verbal and implicit categorization tasks may be due to their failure to collect data during the acquisition phase of the task rather than the testing phase since subjects might be relying on declarative memory (or other) processes to support these later stages of categorization as suggested by Poldrack et al. (2001). In addition, in a system that relies on competition, there might not be differential activation since both systems may attempt to generate a response even though one system consistently wins the competition.

Despite Seger and Cincotta's failure to find evidence directly supporting the COVIS model, their study did identify portions of the striatal system as being active with categorization. And, in fact, the striatal system has been found active with categorization relative to simple response tasks in several imaging studies. For instance, Poldrack and colleagues have adapted a probabilistic-cue categorization task for use in two functional imaging studies and found that the striatal system (the system that is damaged in PARK) is active during probabilistic-cue categorization (Poldrack et al., 2001; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999) with respect to two different baseline conditions. Although the striatal system has long been associated with motor learning, in their first study Poldrack and colleagues (Poldrack et al., 1999) reported the right caudate to be active during acquisition of their categorization task relative to a visually similar baseline condition. This result is consistent with the behavioral deficit observed in early probabilistic categorization learning by PARK subjects. In a second study, Poldrack et al. (Poldrack et al., 2001) addressed in more detail the timing of striatal involvement over the course of probabilistic categorization. Using an event-related design, they concluded that early learning was dependent on hippocampal processes while associative learning systems dependent on the striatal system governed later learning. However, they further noted that the engagement of the medial temporal lobe system was modulated by whether the task encouraged the use of declarative or nondeclarative processes or strategies. In contrast to the hybrid model proposed by Ashby et al. (Ashby et al., 1998) in which competition occurs between categorization systems dependent on frontal and striatal neural systems, Poldrack and colleagues

interpreted the results of their experiment in terms of competition between categorization systems dependent on MTL and striatal neural systems.

The results across functional neuroimaging studies, rather than providing a consensus as to which neural regions of the brain are consistently associated with categorization, actually support the idea that the neural systems supporting categorization may be highly variable and differentially sensitive to manipulations of both the categorization task and the cumulative experience the subject has had with the task. What is beginning to emerge from these studies is that the nature of the categorization task depends on multiple factors and although a neural region – or regions – may be associated with each of these factors, the activity within each of these regions may also change with the cumulative experience the subject has with the categorization task. In addition, the change in neural activation that takes place within these regions over time may occur at different rates for different tasks. At this point, the behavioral, modeling and functional neuroimaging results are still inconclusive as to what neural systems support categorization which would, in turn, provide additional insight into which theory of categorization best reflects the psychological processes supporting categorization behavior. The goal of the experiments presented in Chapters III and IV is to shed additional light on this issue.

CHAPTER III
NEURAL CORRELATES ASSOCIATED WITH
DIFFERENCES IN CATEGORIZATION TASKS

Experiment 1

The fundamental question addressed by the experiments in this dissertation is: What neural processes underlie the explicit learning of perceptual categories? My approach to this question is to use fMRI to investigate experimental paradigms in which the results have been interpreted to support a multiple categorization systems view. I have focused on studying these particular paradigms since they are most likely to yield information about neural regions that may be differentially active with various aspects of categorization. Experiments 1 and 2, presented in this chapter, explore how neural activation may vary with respect to differences between categorization tasks whereas Experiment 3, presented in Chapter IV, explores in greater detail how neural activation associated with a single categorization task may change over the course of category acquisition.

Experiment 1 investigates the neural basis of two categorization paradigms – variations of prototype-distortion and probabilistic-cue classification tasks – that have previously been studied in neuropsychological and normal populations. Although these categorization tasks vary on several dimensions, comparison of the outcomes from behavioral neuropsychological studies utilizing these tasks has been taken as evidence of multiple-categorization systems.

In prototype-distortion tasks, prototypes of stimuli such as dot-patterns, polygons, and objects with multiple binary-valued features are distorted in accordance with an algorithm or rule to create category members (Homa, 1978; Homa & Vosburgh, 1976; Posner et al., 1967; Posner & Keele, 1968; J. M. Reed et al., 1999). These types of stimuli have been used in numerous studies of normal subjects. However, Knowlton and Squire (1993) adapted the prototype-distortion paradigm to test both categorization and recognition in subjects with anterograde amnesia (AA) (Knowlton & Squire, 1993; Squire & Knowlton, 1995). They reasoned that if categorization depended on a nondeclarative memory system and recognition depended on the MTL memory system, then amnesiacs should be impaired on recognition but not categorization relative to normal subjects. In the prototype-distortion paradigm used by Knowlton and Squire,

subjects were shown a series of dot patterns and instructed to point to the center of each pattern when presented. After this exposure phase, subjects were then shown a series of new dot patterns and required to identify them as members (dot patterns generated from the member prototype) or nonmembers (randomly generated dot patterns) of the previously studied category. The category structure used for these dot-pattern categorization paradigms was deterministic in that if a dot pattern was generated from the member prototype, the item was always to be classified as a member of the prototype's category. In the recognition version of this task, subjects were exposed eight times to five dot patterns and were then asked to make old-new recognition decisions with respect to the five old and five new dot patterns. In this study, the results for the AA subjects were comparable to normal controls for categorization but were impaired compared to normal controls for old-new recognition (Knowlton & Squire, 1993) supporting their view that categorization and recognition are dependent on separate memory systems. Squire and Knowlton then followed up their original study by testing a profoundly amnesiac individual, EP, using the same dot-pattern paradigm. While the AA subjects in Knowlton and Squire (1993) had been impaired compared to normal subjects they were still above chance at old-new recognition. By contrast, EP, a profoundly amnesiac subject, was completely at chance on the old-new recognition portion of the task but was still able to perform comparable to normal controls on categorization (Squire & Knowlton, 1995).

The results of these dot-pattern studies of AA populations by Squire and colleagues were followed by several additional studies that extended the results to a larger class of stimuli. Reed and colleagues (J. M. Reed et al., 1999) found results similar to Squire and colleagues when testing amnesiacs using distortions of object-like stimuli (cartoon animals) and Filoteo and colleagues (Filoteo et al., 2001) also reported the ability of amnesiacs to learn and retain a category described by a complex quadratic rule in which the stimuli could be described as distortions of a prototype. Taken together, the results of these studies supported the view that categorization was dependent on a different memory system than recognition and recall since AA subjects showed no impairment on categorization of various classes of prototype-distortion stimuli but were impaired at recognizing or recalling the same types of stimuli and that this dissociation in abilities applied to a broad class of stimuli.

Reber and Squire also used a variation of this prototype-distortion paradigm to test categorization and *recall* in subjects with Parkinson's Disease (PARK) (Reber & Squire, 1999).

When Reber and Squire (1999) tested PARK subjects on their ability to categorize dot-pattern stimuli, they found performance in this neuropsychologically impaired group to be comparable to normal controls (Reber & Squire, 1999). In addition, they found the PARK subject's recall of the testing episode to also be unimpaired compared to the normal controls. However, in contrast to the prototype-distortion studies in which subjects with both AA and PARK were reported to have intact categorization abilities, subjects with PARK had been reported to be impaired compared to AA and normal subjects during the early stages of a probabilistic-cue categorization task (Gluck, Oliver, & Myers, 1996; Knowlton et al., 1996; Knowlton et al., 1994). In this categorization task, the stimuli consisted of a set of four cards. On any trial, subjects were shown one, two, or three of the four possible cards. Each card was probabilistically (rather than deterministically) associated with either the A or B category and on any trial an A or B response could, depending on the assigned probabilities, receive feedback indicating a correct or incorrect response (Knowlton et al., 1994). In testing AA and normal control subjects in this probabilistic task, Knowlton and colleagues (Knowlton et al., 1994) reported that the performance of the AA group was the same as that for controls during the first 50 trials but that the AA group was impaired compared to controls during later learning (trials 100 to 350). Knowlton and colleagues proposed that early probabilistic learning was not dependent on structures supporting LTM but that the normal controls were able to rely on declarative memory for trials during the late learning epochs, thus accounting for their late training advantage over the AA subjects.

In a second probabilistic-cue study, Knowlton and colleagues (Knowlton et al., 1996) used the probabilistic-cue categorization task to again test AA and Control subjects but in this study they also tested a group of subjects with PARK. In this study, the AA and Control subjects performed at comparable levels for trials 41-50 (70.3%, 67.2%) while PARK subjects were comparatively impaired during this same early learning phase (56.3%). However, average percent correct for trials 51-150 revealed no deficits for either the AA or PARK group compared to the normal controls (Control, 66.1%; AA, 59.2%; PARK, 61.4%). The dissociation in performance between AA and PARK subjects during the early epochs of the card categorization task was taken to support the view that early learning during the probabilistic-cue categorization task was dependent on the striatal system – the system damaged in PARK subjects but intact in AA subjects.

The pattern of behavioral results for AA and PARK subjects on the prototype-distortion

and probabilistic-cue tasks relative to each other and to normals has been taken to support a multiple-categorization systems view in which the early stage of the probabilistic-cue categorization task is dependent on a striatal categorization system and prototype-distortion categorization is dependent on a cortical categorization system. However, categorization behavior in neuropsychologically impaired subjects may not be indicative of the neural processes supporting categorization behavior in normal populations. In addition, methodological issues have been identified in the prototype-distortion paradigms used to test amnesiacs that would allow subjects to perform at above-chance levels with significantly impaired memory (Nosofsky & Zaki, 1998) and without any memory whatsoever (Palmeri & Flanery, 1999, 2002).

Since comparison of the results from neuropsychological studies utilizing prototype-distortion and probabilistic-cue tasks has led to the proposal of multiple categorization systems, I chose to directly juxtapose variants of these two tasks in normal subjects while collecting functional images. Comparing the tasks in this way should allow us to identify neural substrates differentially active with the two tasks. Although several neuroimaging studies have examined neural activation in normal subjects during variations of these two tasks (Aizenstein et al., 2000; Poldrack et al., 2001; Poldrack et al., 1999; Reber et al., 2003; Reber et al., 1998a, 1998b; Reber, Wong, & Buxton, 2002; Vogels, Sary, Dupont, & Orban, 2002), none of the studies to date have examined the neural activation associated with both types of tasks within the same group of subjects, and, therefore, it has been difficult to draw conclusions about whether the same or different neural processes support these tasks. In Experiment 1, two groups of subjects were tested in either an early (less than 50 trials) or late (greater than 100 trials) category learning condition. I chose to test subjects in these two conditions since experiments with PARK and AA subjects had indicated that there might be a difference in neural processing between early and late epochs of the probabilistic-cue categorization task. Subjects in each of the learning conditions completed blocks of a prototype-distortion task with a deterministic category feedback structure and blocks of a probabilistic-cue task as well as blocks of control stimuli and blocks of fixation. The only differences between the early and late experimental conditions were the amount of training subjects received before the collection of functional images and the use of new stimulus sets for each imaging run in the early epoch condition.

Although the prototype-distortion paradigms previously used to test AA subjects used an implicit exposure paradigm, the variation of this task used in Experiment 1 was modified to use

an explicit feedback paradigm so it could be tested in conjunction with the probabilistic-cue task. This change in learning condition was necessary since use of an explicit categorization condition in the probabilistic-cue task would almost certainly contaminate subject performance if used at the same time as an implicit learning condition in the prototype-distortion task. This change in the learning condition for the prototype-distortion task from implicit to explicit should, if anything, bias the results against finding a difference in neural activation between the two categorization paradigms since it eliminates a salient difference (implicit vs. explicit learning) between the tasks originally used to demonstrate this dissociation in categorization behavior between AA and PARK subjects. However, this change also eliminates one of the problems inherent with implicit learning paradigms in that normal subjects in these paradigms may actually be aware, or become aware, of the learning task and adopt an explicit learning strategy. The change to an explicit learning paradigm also eliminates a difference in feedback conditions used in the studies of neuropsychological populations. However, differences still remain between the tasks including the type of category structure used – deterministic versus probabilistic – and the type of stimuli used – distortions of a prototype pattern versus a set of four cards with a probabilistic assignment of cues to categories.

If the reported behavioral dissociation between PARK and AA subjects on the prototype-distortion and probabilistic-cue categorization tasks is evidence of neural regions that are differentially sensitive to the differences in these two categorization tasks, I would expect differences in neural activation in normal subjects. In addition, if the probabilistic-cue task is dependent on a striatal learning system, as suggested by the early learning deficits reported for subjects with PARK compared to AA and Control subjects on this task, I would also expect that a subset of these differential activations should be associated with the corticothalamocortical processing loops of the fronto-striatal neural system. Further predictions with respect to neural activation associated with these tasks are less clear. Some studies report no deficit for AA or PARK subjects during later stages of training (Knowlton et al., 1996; Knowlton et al., 1994) suggesting that late categorization is based on a system that is not damaged in either population (the cortico-cortical system) while other studies report a deficit for AA subjects with extended training on the probabilistic-cue task (Gluck et al., 1996) suggesting that performance late in the probabilistic-cue task may be dependent on the MTL system damaged in AA. These changes in reported behavior could be reflected in the data as a differential change in activity with respect to

the task, the training epoch, or as an interaction of task and epoch.

Method

Subjects

Subjects were 22 individuals affiliated with Vanderbilt University who received cash payment or participation credit in an introductory psychology course in exchange for their participation. All subjects were tested individually. Written informed consent was acquired from each individual before testing.

Apparatus and Image Acquisition Parameters

Imaging was performed on a GE 1.5T SIGNA clinical MRI scanner fitted with a high performance local head gradient and RF coils. Functional T2* -weighted images were acquired using a gradient echo, single-shot echo planar imaging sequence (GR-EPI) with a matrix size of 64 x 64, echo time (TE) of 60 ms, flip angle of 90 degrees, TR of 2000 ms, in-plane resolution of 3.75 x 3.75 and slice thickness of 7mm with a 1mm gap. Oblique-axial slices were placed with the bottom slice parallel to the ventral plane of the brain to provide whole brain coverage. Each subject completed three or four functional imaging scans during which both behavioral and imaging data were collected. For anatomical localization, a standard whole-brain, T1-weighted, 3D SPGR sequence was acquired before the functional data collection (124 sagittal slices, in-plane resolution of 0.9375, slice thickness of 1.3mm, no gap).

Stimuli were back-projected onto a screen located outside of the magnet bore which subjects viewed through a periscope mirror attached to the head coil. The mirror was manually adjusted for each subject so that the stimuli would be viewed at the center of the mirror.

Stimuli

Subjects completed blocks of categorization, control, and fixation trials. The stimuli for the categorization and control blocks were either nine-sided polygons (Homa, 1978) (dot-patterns with the dots connected in a pre-defined order) for the prototype-distortion task or combinations of cards with symbols on them for the probabilistic-cue task (Knowlton et al., 1994). One advantage of using polygon stimuli rather than dot pattern stimuli is that the correspondence of the vertices between category members is always known regardless of the level of distortion (Homa, 1978). This correspondence is not always obvious for high distortion dot-pattern stimuli.

Prototype-distortion Condition. The nine-sided polygons used in the prototype-distortion condition (PD) were derived from dot-pattern stimuli. The first step in the stimulus generation process was to generate a random pattern of nine dots on a two-dimensional matrix (Posner & Keele, 1968). To create a coherent novel object the dots were then connected by a solid line to form a nine-sided polygon (Homa, 1984). This polygon then served as a super-prototype from which the category A and B prototypes were derived. The category A and B prototypes were very-high-level distortions of the super-prototype. Category A and B stimuli were then generated in accordance with a computer algorithm by moving the vertices of the A and B prototypes to create new items. The stimuli in the PD condition were presented as yellow lines on a black background.

The underlying category feedback structure for the PD was deterministic (Posner & Keele, 1968). An item generated from the A member prototype was always to be categorized as an A and an item generated from the B prototype was always to be categorized as a B. However, since the category A and B prototypes were generated from the same super-prototype, the items could be confused, and the subject could make errors while learning the category.

Prototype-distortion Control Condition. A second set of polygon stimuli were generated from a second randomly generated prototype to be used in the prototype-distortion control condition (PDX). These stimuli were used as a perceptual control for purposes of the functional imaging analysis and were always presented in white on a black background. Subjects were instructed to attend to the control stimulus and make a button press. No categorization decision was possible since corrective feedback was not provided in this condition.

Probabilistic-cue Condition. Stimuli for the probabilistic-cue condition (PC) were four cards. Each card was identified by a distinctive grouping of squares, circles, triangles, or diamonds. For each trial one, two, or three of the four possible cards would be presented (Knowlton et al., 1994). The cards were displayed as yellow objects and borders on a black background.

The underlying category feedback structure for the PC was probabilistic (Knowlton et al., 1994), and the card combinations were assigned the same outcome probabilities as those used by Knowlton, Squire, and Gluck (1994). These probabilities are listed in Table 2. New sets of PC stimuli for use in the Early condition were generated by randomly assigning the cues to the four cards resulting in new probabilities being assigned to the various cue-combinations.

Probabilistic-cue Control Condition. Stimuli for the probabilistic-cue control condition (PCX) were four cards. Similar to the PC condition, each card was identified by a distinctive design and one, two, or three of the four possible cards would be presented for each control trial. To prevent confusion with the PC stimuli, the cards were presented as white borders and objects on a black background. Subjects were required to attend to the stimulus and press a button. The PCX stimuli were used as a perceptual control for purposes of the functional imaging analysis. No categorization decision was possible on PCX trials since corrective feedback was not provided.

Fixation. Subjects also completed blocks during which they attended to a centrally located fixation cross. Subjects were also required to make a button press during the fixation trials.

Table 2
Probabilistic-cue combinations for card stimuli

Pattern	Cue				P (cue combination)	
	1	2	3	4	Task	P(outcome)
1	0	0	0	1	0.140	0.15
2	0	0	1	0	0.084	0.38
3	0	0	1	1	0.087	0.10
4	0	1	0	0	0.084	0.62
5	0	1	0	1	0.064	0.18
6	0	1	1	0	0.047	0.50
7	0	1	1	1	0.041	0.21
8	1	0	0	0	0.140	0.85
9	1	0	0	1	0.058	0.50
10	1	0	1	0	0.064	0.82
11	1	0	1	1	0.032	0.43
12	1	1	0	0	0.087	0.90
13	1	1	0	1	0.032	0.57
14	1	1	1	0	0.041	0.79

On any trial, 1 of 14 possible combinations of four cues could appear with the probability indicated above (P(cue combination)). Each combination of cues predicted outcome A with the probability P (outcome) shown above and predicted outcome B with a probability of 1-P (outcome). In this example, the cue associations for the A category were 75%, 57%, 43%, and 25% for cues 1 through 4, respectively (Adapted from Knowlton, Squire & Gluck, 1994).

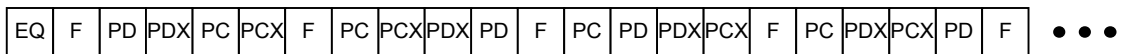


Figure 2. Block Design for Experiment 1. All fMRI runs started with 20 seconds of equalization (EQ) followed by blocks of fixation (F), experimental trials (PD, PC) and control trials (PDX, PCX). Blocks of experimental and control trials were randomized with the constraint that a block of a particular task type could not be repeated until one block of each task had been completed. A block of fixation trials occurred after every four task blocks.

Procedure

This experiment used a mixed design in which one factor – EPOCH – was tested between subjects and a second factor – TASK – was tested within subjects. EPOCH was manipulated by either providing training trials prior to the functional imaging run and maintaining the same set of stimuli during the entire testing session – the Late Epoch (LE) condition – or by withholding all training trials and changing the stimulus set with each functional imaging run – the Early Epoch (EE) condition. TASK included the two experimental conditions previously discussed (PD, PC) and their control conditions (PDX, PCX). Fixation was used as the baseline for purposes of calculating percent signal change from baseline for all conditions.

Procedure – Early Epoch Condition

The purpose of the EE Condition was to test subjects only during the early learning stages, fewer than 50 trials, of both categorization tasks since the primary result reported by Knowlton and colleagues (1994) had been that PARK subjects were impaired compared to AA subjects during this stage of learning for the PC task. This was accomplished by withholding training trials prior to scanning and using new stimuli for each functional imaging run.

Runs. Subjects in the Early Epoch (EE) condition were given written instructions prior to the scan session in addition to a verbal explanation of the experimental procedures. For each fMRI run, a *new* set of PD and PDX stimuli were generated, and a new assignment of probabilities was made to the PC stimuli. Three or four functional imaging runs were collected for each subject depending on time constraints.

Blocks. During each functional imaging run, subjects completed six blocks in each of the four possible task and control conditions: PD, PDX, PC, PCX. Each block included eight trials, and six complete sets of stimulus blocks were collected per run resulting in 48 trials per condition. Blocks of task trials were randomized with the constraint that a block of a particular task type could not be repeated until one block of each task had been completed. A block of fixation trials occurred after every four task blocks.

Trials. Each trial had a total duration of 2500 ms. A fixation cross appeared for 250 ms immediately followed by presentation of the stimulus. Stimulus and response collection was a variable amount of time that was terminated based either on the subject's response or a maximum response time of 1750. If the subject responded prior to the response deadline, corrective feedback was immediately presented and remained on the screen until the end of the

trial (total trial time of 2500 msec). If the subject did not respond prior to the response deadline, response collection was disabled, and corrective feedback was presented for a total of 500 ms (total trial time of 2500 msec).

Procedure – Late Epoch Condition

The purpose of the Late Epoch (LE) Condition was to test subjects only during the "late" learning stages (as defined by Gluck et al., 1996) of the two tasks. This was accomplished by having the subjects complete 100 trials each of PD and PC categorization prior to collection of functional images and by using the same set of stimuli for the entire testing session. Otherwise, all other procedures were the same as in the EE condition.

Late Epoch Pre-training. Subjects completed a pre-training session in which they completed 100 trials each of PD and PC before the collection of functional images. These trials were completed concurrent with acquisition of structural images at the beginning of the scan session. Thus, when subjects in the LE condition started the functional imaging runs, they were effectively in the late learning (greater than 100 trials) phase of both categorization tasks.

Runs. Same as EE condition with the exception that in the LE condition, the stimuli and category feedback structure were **the same** for each run, and each block included only four sets of stimulus blocks per run resulting in 32 trials per run. Four functional imaging runs were collected for each subject in the LE condition. The difference in the length of runs in the EE and LE conditions was due to a scanner upgrade that allowed for longer imaging runs in the EE condition.

Blocks & Trials. Same as EE condition.

Analysis

Behavioral Data Analysis

Behavioral data, percent correct and reaction time (RT) were collected concurrent with the functional imaging runs and analyzed using a standard computer statistics package.

Functional Imaging Analysis

All functional imaging data were analyzed using the AFNI (Analysis of Functional NeuroImages) software package (Cox, 1996).

Individual Subjects. Prior to statistical analysis of the functional images, each subject's time-dependent functional imaging data were examined and corrected for large between phase

fluctuations due to scanner artifacts and subject movement. Images from each functional run were then registered to a single base image to correct for in-plane and out-of-plane subject movement. The six motion parameters used to correct each phase of the imaging run were recorded and used as covariates in the statistical analysis. All images collected during the 20 second equalization period at the beginning of each functional imaging run were discarded. Each subject's structural image was also registered to the same functional base image.

Each subject's motion-corrected functional imaging runs were then reduced to a measure of percent signal change relative to the fixation baseline for each condition. To convert activation to a measure of activity associated with each task, a stimulus function of 1's and 0's representing condition and non-condition time points was created for each of the four task conditions in each run (PD, PDX, PC, PCX) and then convolved with a canonical hemodynamic response function. The four convolved condition functions and the six motion parameters obtained from the image registration process were then submitted as independent variables (IVs), with the voxel activation value as the dependent variable (DV), to a multiple regression analysis. The full model included an estimate of baseline and linear drift in addition to the IVs. The output of this analysis for each individual was then converted to a measure of percent signal change relative to baseline by dividing the beta weights obtained from the regression analysis for each condition in each voxel by the beta weight for the baseline condition in each voxel. The results of this analysis were then placed into the standard coordinate system of Talairach and Tournoux (Talairach & Tournoux, 1988).

Group Analysis. Since the inflation of Type I error is a concern in fMRI analyses, I chose to limit the statistical analysis to neural regions that showed maximal activity for the categorization relative to the control conditions. To identify these regions-of-interest (ROIs), the percent signal change values for each subject in both the EE and LE conditions were submitted to an ANOVA to determine voxels in which there was a significant main effect of categorization (PD, PC) relative to control (PDX, PCX). Regions of maximum intensity within this map were then identified and spherical ROIs with a 4 mm radius (257 vmul) were created at these voxel coordinates for use in additional statistical analysis of the data. The advantage of using a smaller ROI defined in the manner described above is that it provides better localization of the activation as well as avoids dilution of the measured neural activation that can result from averaging results across a large number of voxels with varying degrees of response to the experimental conditions.

The next step in the ROI analysis was to calculate a measure of *net* percent signal change for use in a second analysis limited to the ROIs showing a significant main effect of categorization. The values used in this ROI-based analysis were the net percent signal change for each experimental condition derived by subtracting the percent signal change value for each of the control conditions from the percent signal change value for each of the experimental task conditions. This calculation resulted in four IVs as follows: EPD (Early PD – Early PDX), LPD (Late PD – Late PDX), EPC (Early PC – Early PCX), and LPC (Late PC – Late PCX). The average net percent signal change reported for each of the ROIs was then investigated using a 2x2 mixed-effect ANOVA.

Results

Behavioral Data

As illustrated in Figure 3, subjects were able to acquire information about category membership in both the prototype-distortion and probabilistic-cue tasks. Consistent with the method used in other studies using probabilistic-cue tasks (Gluck et al., 1996; Knowlton et al., 1996; Knowlton et al., 1994), trials in the PC task were scored based on the category for which they had the highest probability of membership. That is, if a card combination was included in the A category 75% of the time, a response of ‘A’ by the subject was scored as a correct response

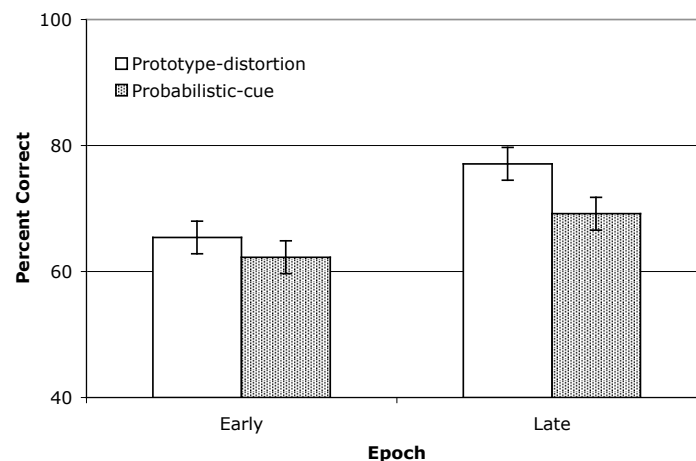


Figure 3. Mean percent correct in the four categorization conditions of Experiment 1. In the early condition, mean percent correct was 65% for the Prototype-distortion task and 62% for the Probabilistic-cue task. In the late condition, mean percent correct was 77% for the Prototype-distortion task and 69% for the Probabilistic-cue task. Error bars are *rmse*.

regardless of the feedback received. Items with a 50% chance of being included in either category were excluded from the accuracy analysis. One subject in the late condition was excluded from both the behavioral and functional imaging analysis as s/he failed to achieve above-chance performance on either categorization task.

A mixed-effect ANOVA for the average percent correct for each subject by condition confirmed a significant main effect of EPOCH [$F(1,19) = 13.281, p < .05$]. Neither the main effect of TASK [$F(1,19) = 4.332, p > .05$] nor the interaction of EPOCH and TASK [$F(1,19) = 0.817, p > .05$] was significant. A mixed-effect ANOVA for the average reaction time (RT) for each subject confirmed a significant main effect of TASK [$F(1,19) = 7.28, p < .05$] – RTs were greater for the PC relative to the PD task. Neither the main effect of EPOCH [$F(1,19) = 0.929, p > .05$] nor the interaction of EPOCH and TASK [$F(1,19) = 0.018, p > .05$] was significant. Timeouts (failure to respond before the response deadline) averaged less than 2%.

Neuroimaging Data

Categorization Regions-of-Interest (ROIs). Table 3 lists the voxel coordinates (Talairach & Tournoux, 1988) for the maximum intensity voxel of the 20 regions-of-interest in which significant changes were observed for the categorization tasks relative to the control tasks. These coordinates then served as the center of mass for the spherical ROIs created and used for the ROI-based statistical analysis of the functional images as discussed in the analysis section.

After identification of the ROIs, the four *net* activation values (EPD, EPC, LPD, LPC) for the 20 ROIs were further investigated in a two-factor mixed-effect ANOVA. As reported in Table 3, seven of these neural regions (4 frontal, 2 occipital, 1 parietal) were found to have a significant main effect of TASK and one region in the right inferior parietal lobe was found to have a significant interaction of EPOCH and TASK. None of the regions were found to have a significant main effect of EPOCH.

Discussion

The differences between the two experimental conditions tested in Experiment 1 were broad - as were the task differences in the behavioral neuropsychological studies that motivated the study - and therefore, the conclusions that can be drawn from these results are somewhat limited. Although the behavioral results indicate that performance on the prototype-distortion

Table 3.
Experiment 1 Regions of Interest and ANOVA Results

LR	AP	IS	Hem	Brain Region	BA	Effect	<i>F</i>	<i>p</i>
<u>Frontal</u>								
0	57	0	Medial	Medial Frontal Gyrus	10			
37	46	17	Right	Middle Frontal Gyrus	10	TASK	8.080	0.010
-21	31	42	Left	Middle Frontal Gyrus	8	TASK	7.155	0.015
-29	22	-4	Left	Inferior Frontal Gyrus	47			
32	22	-4	Right	Inferior Frontal Gyrus	47			
1	14	48	Medial	Superior Frontal Gyrus	6	TASK	5.573	0.029
47	10	32	Right	Middle Frontal Gyrus	9			
0	-17	43	Medial	Paracentral Lobule	31	TASK	6.649	0.018
<u>Limbic</u>								
8	2	9	Right	Caudate		Body		
-8	-2	11	Left	Thalamus		*		
<u>Temporal</u>								
-45	-58	31	Left	Superior Temporal Gyrus	39			
<u>Parietal</u>								
36	-39	45	Right	Inferior Parietal Lobule	40	EPOCH X TASK	9.254	0.007
0	-46	34	Medial	Precuneus	31	.	.	.
28	-51	41	Right	Superior Parietal Lobule	7	.	.	.
-32	-53	45	Left	Inferior Parietal Lobule	40	TASK	7.015	0.016
24	-63	50	Right	Superior Parietal Lobule	7			
-48	-66	33	Left	Angular Gyrus	39			
<u>Occipital</u>								
37	-76	7	Right	Middle Occipital Gyrus	19	TASK	5.823	0.026
-40	-78	3	Left	Middle Occipital Gyrus	19	TASK	7.976	0.011
-10	-89	4	Left	Lingual Gyrus	17			

Regions of Interest based on maximum *F* values for categorization conditions relative to control conditions. R = Right (+), L = Left (-); A = Anterior(+), P = Posterior(-), S = Superior (+), I = Inferior (-); Hem = Hemisphere; Brain Regions = Talairach Label; BA = Brodmann Area; * no Brodmann label. Effects = ROIs with significant effects as confirmed by ANOVA; TASK = Main Effect of Task; EPOCH X TASK = Interaction of Epoch and Task; *F* = *F* value for reported effect from ANOVA with degrees of freedom = 1,19; *p* = achieved *p*-value for reported effect from ANOVA.

and probabilistic-cue tasks was roughly equated, seven neural regions exhibited differential activity with performance of the PD relative to the PC task. This difference in neural activation in the absence of significant differences in behavioral performance would tend to support the conclusion that different categorization systems support these two categorization tasks or that different areas are engaged in certain kinds of tasks². In addition, although I did not find

² Although differential activation in fMRI may be a necessary condition to conclude that a separate system exists, differential activation by itself is not sufficient to conclude that a separate system exists. fMRI of normal subjects,

significant differences associated with task in the caudate or thalamus, as might be predicted from previous fMRI studies, four of the activations (three frontal and one cingulate) were associated with neural regions that are part of the extended fronto-striatal processing system which, as discussed in Chapter II, innervates dorsal lateral prefrontal cortex (DLPFC), medial orbitofrontal cortex (MOC), and anterior cingulate cortex (ACC) via the corticostriatohalamocortical loops.

The three frontal regions differentially active with respect to the two categorization tasks (PC > PD) have previously been associated with a variety of tasks requiring working memory (Cabeza & Nyberg, 2000). In particular, the superior gyrus (BA 6) has been associated more generally with verbal working memory tasks (Fiez, 1996; Coull, 1996) whereas the right middle frontal gyrus (BA 10) has been associated with numerous problem solving tasks such as the Wisconsin Card Sorting task (Goldberg, 1998; Berman, 1995), the Tower of London task (Baker, 1996), and Raven's Analytical task (Prabhakar, 1997) as well as a conceptual reasoning task (Rao, 1997). Activation of the left middle frontal gyrus (BA 8) has been associated with language (Bookheimer, 1995; Bottini, 1994), semantic retrieval associated with categorization (Demb, 1995; Demonet, 1994) and generation tasks (Phelps, 1997), and conceptual priming (Demb, 1995; Gabrieli, 1996). This main effect of TASK may indicate that the categorization processes supporting probabilistic-cue categorization rely upon holding and manipulating inputs to the categorization process in a way that is different than the categorization processes supporting prototype-distortion categorization since activation was greater for the probabilistic-cue task relative to the prototype-distortion task in each of these frontal regions. These results would also appear to be consistent with those of a recent study in which performance on a probabilistic-cue classification learning task was facilitated by electrical stimulation of the prefrontal, but not occipital, cortex (Kincses, et al., 2003) supporting the finding that the neural locus of process differences in early epochs of PD and PC categorization is associated with cognitive functions supported by the frontal lobes rather than the basal ganglia. In addition, a recent study by Gluck and colleagues (Gluck, Shohamy, & Myers, 2002) of the strategies used by normal subjects in solving the probabilistic-cue task reports that normal subjects use one of several strategies, including two simple rule-based strategies, when solving certain probabilistic-

by itself, is not sufficient to make conclusions with respect to multiple-systems. See Chapter V and Price and Friston (2003) for a more detailed discussion of this issue.

cue tasks. The results of this study (Gluck et al., 2002) taken with the pattern of differential neural activation reported in the current experiment suggest that the deficits reported for PARK subjects during the early stages of probabilistic-cue classification may be due to an impaired ability to apply rules that is dependent on intact frontal lobe function rather than an impaired ability to make a stimulus-response association that is dependent on intact basal ganglia functions. The possibility of a hybrid categorization system that includes an early, rule-based categorization system is examined in greater detail in Experiment 3, presented in Chapter IV.

An interaction of TASK and EPOCH was reported for one region in the right inferior parietal lobule (IPL). Activity in parietal areas has been identified in several categorization studies (Aizenstein et al., 2000; Reber et al., 2002; Seger et al., 2000; Vogels et al., 2002) however no single factor across these studies appears to be exclusively associated with these activations. One possibility for this pattern of activation is that the different stimuli used for the two experimental tasks are processed in different ways. For instance, the prototype-distortion task uses a single, nine-sided polygon for each trial whereas the probabilistic-cue task uses one to three cards for each trial. Figure 4 shows that activation in the right IPL decreased between the Early and Late epochs of the prototype-distortion task whereas activity for the probabilistic-cue task increased from the Early to Late epoch. Shafritz, Gore and Marois (2002) have reported that regions of superior and inferior parietal cortex are more active in feature conjunction tasks – like the probabilistic-cue task – than in single-object tasks – like the prototype-distortion task

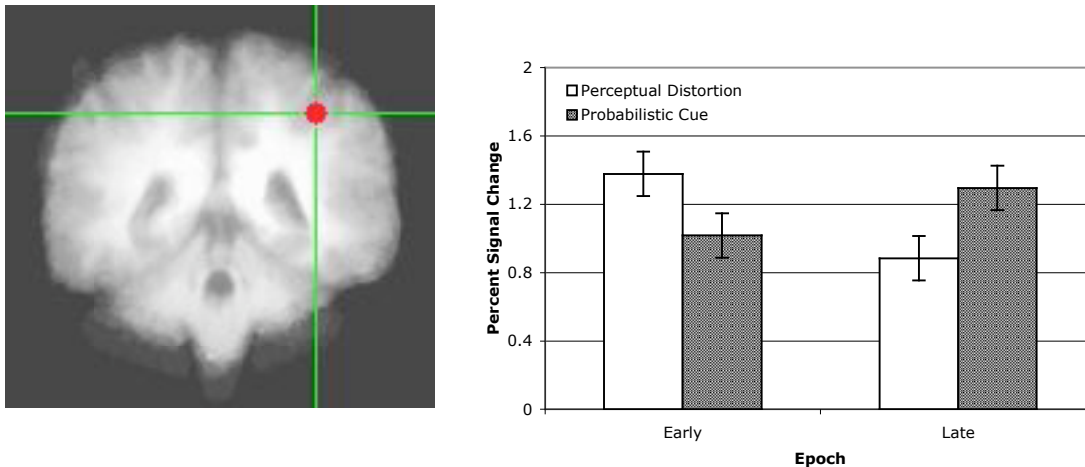


Figure 4. EPOCHxTASK interaction – right inferior parietal lobule. Panel on the left shows the location of the ROI in a coronal plane. Relative activation for each experimental condition (net of control) is presented in the left panel. Structural image is an average image from all subjects. Error bars are *rmse*.

(Shafritz, Gore, & Marois, 2002) and therefore, in the current experiment, it may be the relationship between the type of stimuli and the subject's strategy for learning the categories that resulted in the reported interaction. One can speculate that a single card in the probabilistic-cue task is a simple stimulus compared to a single ten-sided polygon in the prototype-distortion task. If subjects adopt a strategy during early epochs of the probabilistic-cue task of attending to a single card, as suggested by Gluck et al. (2002) and then change to a multi-cue strategy in late epochs of the probabilistic-cue task in which they attend to all of the presented cards, the pattern of increasing neural activation between Epochs of the PC task shown in Figure 4 might be expected. In the case of the single-object stimuli used in the PD task, this switch in strategy may not take place and the resultant decrease in activation between the Early and Late epoch may simply reflect the general reduction in activation that occurs with task repetition.

One unanticipated result of this study was that none of the ROIs tested were associated with a significant main effect of EPOCH. Given the significant effect of EPOCH reported for the behavioral data and previous research reporting a dissociation in the pattern of results reported for early and late epochs of the PD and PC tasks for AA, PARK, and normal subjects, one might have also predicted a difference in neural activation with respect to this factor. However, whereas TASK was tested within subjects, EPOCH was tested between subjects and therefore, the failure to find a result in the between subject factor may simply be due to the limited power to identify an effect related to this aspect of the experimental design.

As previously noted, although differential results were reported for the two experimental tasks in Experiment 1, there are problems with interpretation due to the choice of paradigm that was dictated by the extant behavioral neuropsychological studies. First, as previously noted, the experimental tasks varied on more than one factor. The tasks differed in the type of stimuli used, polygon versus cards; the feedback structure assigned to the stimuli, deterministic versus probabilistic; and the level of variation in the stimuli, multiple distortions of a prototype versus a single set of four cards; therefore, any differential activation associated with the tasks could be attributed to one or more of these factors. A second limitation of Experiment 1 was that both tasks used an explicit learning paradigm. Several of the behavioral neuropsychological studies that motivated the current work used implicit learning paradigms and therefore, differences noted between performance by neuropsychological populations and the normal subjects tested in Experiment 1 could be due to the differences in learning and testing associated with using an

implicit or explicit learning paradigm. Experiment 2, below, will address some of the limitations in interpretation due to the experimental design used in Experiment 1 and the categorization fMRI literature in general whereas Experiment 3, presented in Chapter IV, will address changes in neural activation over the course of category acquisition. Addressing differences between implicit and explicit categorization paradigms is beyond the scope of this dissertation.

Experiment 2

The results from Experiment 1 are generally consistent with the inference that has been drawn in the behavioral neuropsychological literature that there is a difference in the neural systems underlying prototype-distortion and probabilistic-cue categorization tasks since there were significant main effects of TASK reported for a number of neural regions associated with the fronto-striatal learning and memory system. However, since more than one experimental factor differed between the two experimental conditions of Experiment 1, no definitive conclusions could be drawn with respect to the factor or factors directly associated with these results. Therefore, one of the goals of Experiment 2 is to address some of the limitations of Experiment 1.

A second goal of Experiment 2 is to address some of the limitations found with respect to various categorization tasks reported in the literature. As noted in Chapter II, activation with respect to categorization relative to control has been reported in the MTL, cortico-cortical, and fronto-striatal learning and memory systems. However, these results are not consistently found between studies. Although some studies have specifically addressed issues with respect to multiple-categorization systems (Aizenstein et al., 2000; Patalano et al., 2001; Poldrack et al., 2001; Poldrack et al., 1999; Reber et al., 2003), the results across the limited number of neuroimaging studies of categorization published to date have failed to reach any clear consensus with respect to specific brain regions subserving any particular categorization system or task. One reason for this failure to achieve consensus across studies may be due to differences in the experimental designs used to test subjects. In adapting categorization tasks for fMRI studies, several aspects of the categorization task structure are consistent. For example, in a typical explicit categorization paradigm, stimuli are assigned to two or more categories according to a method determined by the experimenter. Subjects then participate in some type of learning task in order to acquire the object-category assignments (category training) and are then tested either

on how well they have learned the category assignments (category testing) or on how they apply their category knowledge to new items (category transfer). However, despite these consistencies in the structure of the categorization experiments, there are still many aspects of the experimental paradigms that vary. For instance, in many experiments the stimuli are assigned to the category deterministically (object X is *always* an A) but other studies use probabilistic category assignments (object X is *usually* an A, but sometimes a B). Similarly, subjects can be tested on category transfer after training that is either relatively short (40 training trials) or long (256 training trials). Finally, subjects often learn the category structure via an *explicit feedback* paradigm in which they see a stimulus, offer a response, and then receive corrective feedback (often referred to as supervised, intentional, or explicit learning). Alternatively, subjects can learn the category structure via an *explicit exposure* paradigm in which they see a stimulus and are told its category membership – similar to a paired associates task – or via *implicit (or incidental) exposure* paradigms in which they see a series of items from a single category without being told before the exposure task of any relationship between the stimuli being viewed and any subsequent categorization tests. Given the number of possible task differences and the relatively small number of categorization studies utilizing fMRI it is not surprising that there has been little consistency in the experimental conditions tested or the results reported from these studies.

A review of the experimental conditions in fMRI studies of categorization with paradigms similar to the paradigms used in this dissertation (Aizenstein et al., 2000; Patalano et al., 2001; Poldrack et al., 2001; Poldrack et al., 1999; Reber et al., 1998a, 1998b; Reber et al., 2002; Seger & Cincotta, 2002; Seger et al., 2000; Vogels et al., 2002) reveals that in these studies functional images were gathered during *both* the learning and testing or transfer stages of the task in only 3 of the 10 studies. Of the remaining 7 studies, functional images were collected only during the transfer portion of the task and therefore evaluation of changes in activity with respect to early category acquisition is not possible in these studies. Similarly, 8 of the 10 studies used deterministic category feedback structures while the remaining two studies utilized probabilistic feedback structures. And, in these studies, 4 used an implicit learning paradigm and 6 used an explicit learning paradigm. In addition to the differences in the categorization paradigm, the stimuli used in these categorization studies have also varied. In general, the probabilistic-cue studies have used a variation of the card categorization task described in

Experiment 1 while the deterministic studies have used a variety of stimulus sets such as dot patterns, polygons, checkerboard patterns, and simple objects with variable numbers of binary features. The stimulus sets used in the deterministic studies were all generated based on distortions or variations from a category prototype. Other differences that have not been controlled for across studies include level of performance and the baseline conditions used to evaluate the experimental conditions.

Although there is certainly nothing incorrect in studying a limited set of conditions within an experiment, this approach should also limit the conclusions that can be drawn across studies. Given that one of the impediments to understanding the fMRI literature as it relates to categorization is that it is difficult to draw conclusions between studies that vary by more than one study parameter, one of the goals of Experiment 2 is to study a limited set of parameters that have been associated with differential activation in a variety of studies within one group of subjects. Therefore, in Experiment 2, I used a within-subjects, 2 (learning versus testing) x 2 (deterministic versus probabilistic category structure) x 2 (exposure versus feedback learning) factorial design to study neural activation associated with three factors that had not previously been studied in a single subject group. Designing Experiment 2 in this way also achieves a second goal, which is to partially replicate the Early Epoch condition of Experiment 1 using a single stimulus set – a difference in conditions that had limited interpretation of the results of Experiment 1. I chose to further investigate the Early, rather than Late, learning condition since differences in performance between neuropsychological populations had been consistently associated with early categorization. Most important, the paradigm used in Experiment 2 will provide additional insight into whether the differences reported in the behavioral neuropsychological literature and previous fMRI studies with respect to the prototype-distortion and probabilistic-cue task are associated with the underlying category structure (deterministic versus probabilistic) or the type of learning paradigm used to acquire the category (exposure versus feedback) – two factors that had been confounded in previous studies (Poldrack et al., 2001; Poldrack et al., 1999) as well as in drawing conclusions between studies (Knowlton et al., 1996; Knowlton & Squire, 1993; Knowlton et al., 1994; Squire & Knowlton, 1995).

In Experiment 2, I collected functional images while subjects participated in the early stages of category learning and testing. During one session, subjects completed learning and transfer tasks with stimuli that had a deterministic category structure. During another scan

Table 4
Experimental Conditions of Experiment 2

LEARNING				TESTING
<u>Exposure</u>		<u>Feedback</u>		
<u>Deterministic</u>	<u>Probabilistic</u>	<u>Deterministic</u>	<u>Probabilistic</u>	
DEL				DET
	PEL			PET
		DFL		DFT
			PFL	PFT

Two learning (L) conditions: Feedback (F) or Exposure (E) with two possible category structures: Deterministic (D) or Probabilistic (P) which then result in four possible testing (T) conditions based on the type of learning condition.

session, the same group of subjects completed learning and testing tasks with stimuli that had a probabilistic category structure. During each session, subjects completed four pairs of functional imaging runs (8 runs per session). For each pair of runs, the first run was always a category learning run during which the subject learned the category via either *feedback* category learning or *exposure* category learning. Following each training run, the subject was then tested without feedback. The data for learning and testing were collected in separate functional imaging runs (rather than in alternating blocks of learning and testing within a single run) to avoid possible carryover effects between the learning and transfer tasks (Braver, Reynolds, & Donaldson, 2003; Visser et al., 2003).

The stimuli were similar to the card stimuli used in Experiment 1 however, the set of stimuli was expanded so new items could be used for each pair of imaging runs (See Appendix A). For each complete scanning session, subjects were to be tested twice in each of the different conditions using one of the two types of category feedback structures (deterministic or probabilistic). The category structure used in the first scanning session was randomly assigned for each subject and subjects were then tested in the second day's scanning session using whichever category feedback structure was not tested in the first day's session. The eight conditions tested in Experiment 2 are summarized in Table 4.

Functional imaging of these conditions will allow us to systematically evaluate the neural basis of different types of learning (feedback versus exposure) and category feedback structure (probabilistic versus deterministic) at both the time of initial learning and at the time of category testing for each of the possible learning conditions.

Method

Subjects

Subjects were 10 individuals affiliated with Vanderbilt University who received cash payment for their participation. All subjects were tested individually and self-reported to be right-handed with normal or corrected-to-normal vision. Written informed consent was acquired from each individual before each testing session.

Apparatus and Image Acquisition Parameters

Imaging was performed on a GE 3T SIGNA research MRI scanner fitted with a high performance local head gradient and RF coils. Functional T2*-weighted images were acquired using a gradient echo, single-shot echo planar imaging sequence (GR-EPI) with a matrix size of 64 x 64 (FOV 240mm), echo time of 60 ms, flip angle of 70 degrees, TR of 2500 ms, in-plane resolution of 3.75 x 3.75, and slice thickness of 7mm, with 0 mm gap. 18 axial slices were placed starting at the ventral surface of the temporal lobe and extending in the superior plane to provide whole brain coverage. Subjects completed eight functional imaging scans during each of two imaging sessions during which both behavioral and imaging data were collected.

For anatomical localization, a standard whole-brain, T1-weighted, 3D FSPGR sequence was acquired after functional data collection (124 sagittal slices, in-plane resolution of 0.9375, slice thickness of 1.2 mm, no gap).

Stimuli

Stimuli for each condition were four cards, similar to the cards used for the probabilistic-cue task in Experiment 1 and other probabilistic-cue experiments reported by Knowlton and colleagues (Gluck et al., 1996; Knowlton et al., 1996; Knowlton et al., 1994). However, for purposes of this experiment, an expanded set of 64 card stimuli were created so that a unique set of four cards could be used for each set of learning/testing runs. The 64 card stimuli were constructed from eight nameable objects arranged in eight different configurations. For any set of four stimuli presented to a subject, the same object/configuration pairing was not used for more than one set of learning/testing functional imaging runs. Subject to this constraint, all stimuli were randomly assigned to conditions and subjects. Stimuli were presented using a mirror focused on a screen outside of the scanner on which the stimuli were projected.

Probabilistic Category Structure. For each trial one, two, or three of the four possible cards were presented in one of 12 stimulus combinations. The cards were displayed as blue

objects and borders on a white background. The underlying category feedback structure for the probabilistic condition was similar to the structure used by Knowlton and colleagues with the exception that the two stimuli in the Knowlton and Squire paradigm that had a 50% probability of being a member of the A or B category were excluded from the set. Stimuli for each pair of learning/testing imaging runs were assigned by drawing sets of four cards from the 64 card set without replacement and randomly assigning the cues to the cards.

Deterministic Category Structure. The deterministic stimuli were constructed in the same manner as the probabilistic stimuli. However, identification of the item as a member of category A or B was deterministic. For instance, if an item in the deterministic condition was assigned a cue pattern that would result in an 'A' category assignment 75% of the time in the probabilistic condition, the item would be treated as an 'A' 100% of the time in the deterministic condition. Deterministic stimuli were presented with the same frequencies as the probabilistic stimuli. Pilot testing confirmed that the category feedback assignments of the four training conditions resulted in equated performance (percent correct) in the four transfer conditions.

Control Conditions. Stimuli for the Control conditions (probabilistic and deterministic) were drawn from the same pool of 64 cards. However, to distinguish the control stimuli from the categorical stimuli, the control stimuli were presented as red objects and borders on a white background. As with the categorical stimuli, one, two or three of the four possible cards were presented for each control trial and subjects were simply required to press a button within a response deadline. Category feedback or identification was never provided for these items. Presentation of control stimuli was yoked to presentation of categorization stimuli to equate the experimental and control conditions on exposure. Thus, for each block of 8 categorization trials (learning or testing) a block of 8 control trials was also presented.

Fixation. Subjects also completed three 20-second periods of fixation presented at the beginning, midpoint, and end of each imaging run. During fixation subjects attended to a centrally located fixation cross presented in black type on a white background.

Procedure

Functional Imaging Sessions. Subjects participated in two functional imaging sessions. Eight functional imaging runs (two per condition) were collected during each session for a total of 16 functional imaging runs per subject over the two scanning sessions. On the first day of scanning subjects learned the categorical assignments using either a deterministic or probabilistic

category structure. The category structure for session 1 was determined at random and the remaining, unused, category structure was then used approximately one week later for session 2. The category structures were blocked by session in order to minimize subject confusion during the learning and testing sessions. Similarly, the order of presentation for the two types of learning runs (exposure or feedback) was randomly assigned to the four possible learning runs during each session (runs 1,3,5, and 7) subject to the constraint that two runs of exposure and two runs of feedback learning were assigned for each scanning session. Scan numbers 2, 4, 6, and 8 in each scanning session were always transfer runs which did not vary from each other except for the type of learning paradigm that had been used in the immediately preceding learning run and the card stimuli assigned to the condition.

Functional Imaging Runs. Stimuli in the learning and testing runs were presented in a series of 8-trial blocks. For the learning scans, each scan included 6 blocks of learning trials interleaved with 6 blocks of control trials in an ABBA (Petersson, Elfgren, & Ingvar, 1999; Skudlarski, Constable, & Gore, 1999) design to control for linear drift during the scan. Testing scans were structured the same as the learning scans with the only difference being the absence of corrective feedback.

Trials. All trials had the same temporal structure. Each trial started with a 250 ms fixation cross, a stimulus was then presented for 1750 ms during which the subject could respond. If the subject responded before the 1750 ms deadline, the stimulus would remain on the screen and the computer would provide an appropriate response to the subject's key press. If the subject failed to respond within 1750 ms, the stimulus would remain on the screen but the computer would then cancel the response collection process and instead would provide an appropriate response. Trials terminated after a total elapsed time of 2500 ms per trial. Table 5 outlines the response request and feedback screens for each condition. All responses were made using a button box held in the subject's right hand during the scanning session. The index finger was used to record an 'A' response and the middle finger was used to record a 'B' response. Similar trial timing parameters were used in Experiment 1 and results from that experiment indicated that subjects were able to learn the categories within these timing constraints. Subjects received complete written instructions via a self-paced computer presentation as well as verbal instructions before collection of functional images.

TABLE 5.
Structure of Learning, Testing, and Control Trials in Experiment 2

<u>Duration</u>	<u>Feedback Learning Trials</u>	<u>Exposure Learning Trials</u>	<u>Testing Trials</u>	<u>Control Trials</u>
Fixation = 250 ms	+	+	+	+
Response Request ≤ 1750 ms	A or B? (Subjects press index finger for an A, middle finger for B)	Category A (Subjects press index finger to acknowledge A, middle finger for B)	A or B?	Press a Key.
Correct Response ≥ 500 ms	Correct, Category A	Correct, Category A	Response Collected.	Response Collected.
Incorrect Response ≥ 500 ms	Incorrect, Category A	Incorrect, Category A	Response Collected.	Response Collected.
No Response = 500 ms	Response not Detected. Category A	Response not Detected. Category A	Response not Detected.	Response not Detected.

First column describes the time allocations for each 2,500 ms trial. Columns 2 through 5 describe the possible actions and screen displays for each stage in a trial for the four possible trial types. The stimulus appeared on the screen immediately after the fixation cross and stayed on the screen until the end of the trial (2,250 ms). ms = milliseconds

Analysis

Behavioral Data Analysis

Behavioral data (percent correct and RT) were collected concurrent with the functional imaging runs and analyzed using a standard computer statistics package.

Functional Imaging Analysis

All functional imaging data were analyzed using the AFNI (Analysis of Functional NeuroImages) software package (Cox, 1996).

Individual Subjects. After collection of functional images, the following steps were taken to convert each functional imaging run into a measure of percent signal change for each experimental variable. First, an algorithm was used to identify outliers (spikes) in each series of images and, in the case where an outlier was identified, replace the outlier with a new value based on the average of images immediately prior to and after the outlier. Next, to correct for small amounts of motion, each functional image was registered to a base functional image

acquired immediately before acquisition of the 3-dimensional structural MRI. The six motion parameters required to register each image to the base image were recorded for use in the subsequent regression analysis. Next, a mean activation value was determined for each voxel within each functional imaging run and then each series of functional images was normalized by dividing the activation in each voxel at each point in time by the mean value for that voxel for that imaging run and dividing the result by 100. The resulting activation value was then spatially smoothed using a 4mm gaussian filter. Finally, the functional activation profiles were submitted to a regression analysis that included the 6 motion parameters generated during volume registration as well as two ideal functions - one representing the hypothetical hemodynamic response function for neural regions active with respect to the blocks of control stimuli and one representing the hypothetical hemodynamic response function for neural regions active with respect to the blocks of categorical stimuli. The results of this regression analysis were then converted to a measure of percent signal change relative to the fixation baseline. Each subject's data were then placed into the coordinate system of Talairach and Tournoux (Talairach & Tournoux, 1988) for inclusion in the group level analysis.

Group Analysis. As in Experiment 1, to control for Type I error rate I chose to limit the statistical analysis to a limited number of neural regions that showed maximal activity for categorization relative to control. To identify these ROIs, the percent signal change values for each subject in the experimental and control conditions were submitted to an ANOVA to determine voxels in which there was a significant main effect of categorization relative to control. Regions of maximum intensity within this map were then identified and spherical ROIs (257 vmul – approximately 2.5 voxels in original acquisition matrix) were created at these voxel coordinates for use in additional statistical analysis of the data. The advantage of using a smaller ROI defined in the manner described above is that it provides better localization of the activation as well as avoids dilution of the measured neural activation that can result from averaging results across large numbers of voxels with varying degrees of response to the experimental conditions. The next step in the ROI analysis was to calculate a measure of *net* percent signal change for use in the ANOVA by subtracting the percent signal change value for each of the experimental control conditions from the percent signal change value for each of the experimental task conditions. This subtraction resulted in net activation values for the eight IVs listed in Table 4. The average

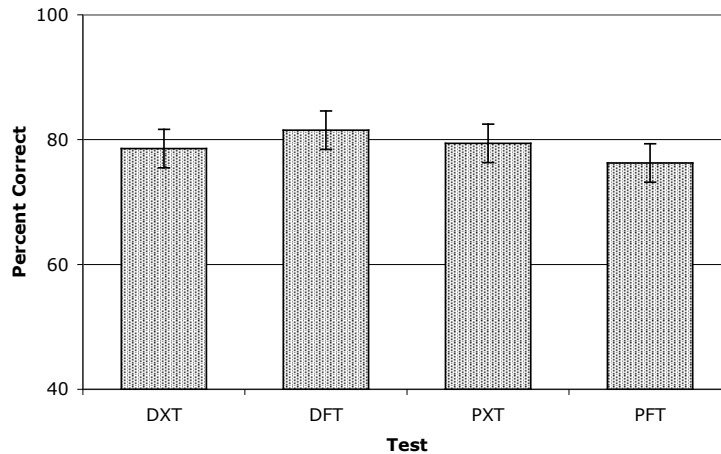


Figure 5. Performance in the four test conditions of Experiment 2. The only difference between test conditions was the learning condition used for category acquisition *prior* to the test runs. There was no significant main effect of category structure (deterministic versus probabilistic: ANOVA [$F(1,8)=0.417$]) or type of learning (feedback versus exposure: ANOVA [$F(1,9)=0.082$]). DXT = deterministic exposure; DFT = deterministic feedback; PXT = probabilistic exposure; PFT = probabilistic feedback. Error bars are *rmse*.

neural activation reported for each of the ROIs was then investigated using a 2 (Learning/Testing) x 2 (Feedback/Exposure) x 2 (Deterministic/Probabilistic) ANOVA.

Results & Discussion

Behavioral

Since the task during exposure training consisted of simply acknowledging the correct category by making a button press, subjects were expected, and did, perform at ceiling. Average percent correct in the two feedback learning conditions, DFL and PFL, was 71% and 62%, respectively. Although it was not possible, due to the structure of the different learning tasks, to equate performance between the four learning conditions, Figure 5 illustrates that category *testing* after equal numbers of learning trials in these four conditions was equated across the four conditions as confirmed by a two-way ANOVA. Similar to the results for percent correct, analysis of RTs revealed no significant main effects related to the three experimental factors (all F 's < .367, all p 's > .05). Timeouts accounted for less than 2 percent of test trials.

Table 6.
Experiment 2 Regions of Interest and ANOVA Results

LR	AP	IS	Hem	Brain Region	BA	Effect	F	p
<u>Frontal</u>								
46	13	40	Right	Middle Frontal Gyrus	8	LTxFX	11.76	.001
-28	4	56	Left	Middle Frontal Gyrus	6	FX	5.714	.020
2	58	-4	Right	Medial Frontal Gyrus	10			
50	15	0	Right	Inferior Frontal Gyrus	47			
-39	14	-5	Left	Inferior Frontal Gyrus	47	LTxFX	5.255	.026
<u>Limbic</u>								
1	18	43	Right	Cingulate Gyrus	32	LTxFX	8.106	.006
8	-4	14	Right	Thalamus	AN	FX	4.019	.050
-12	-14	16	Left	Thalamus	*	LTxFX	14.52	.000
43	-12	10	Right	Insula	13			
<u>Temporal</u>								
-51	-60	29	Left	Middle Temporal Gyrus	39	LT	14.66	.000
-51	-60	29	Left	Middle Temporal Gyrus	39	FX	9.830	.003
47	-62	-3	Right	Inferior Temporal Gyrus	19			
<u>Occipital</u>								
-34	-70	-8	Left	Lingual Gyrus	18	LTxFX	4.520	.038
-7	-80	-8	Left	Lingual Gyrus	18	LTxFX	6.938	.011
37	-67	-6	Right	Lingual Gyrus	18			
-8	-74	17	Left	Cuneus	18	LTxFX	7.279	.009
<u>Parietal</u>								
3	-61	60	Right	Precuneus	7			
-3	-47	36	Left	Precuneus	31			

Regions of Interest based on maximum F values for categorization conditions relative to control conditions. R = Right (+), L = Left (-); A = Anterior(+), P = Posterior(-), S = Superior (+), I = Inferior (-); Hem = Hemisphere; Brain Regions = Talairach Label; BA = Brodmann Area; * no Brodmann label. Significant Effects = ROIs with significant effects confirmed by ANOVA; LT = Main Effect of Learn vs Test; FX = Main Effect of Feedback vs. Exposure; LTxFX = Interaction of Learn vs. Test and Feedback vs Exposure; F = F value for reported effect from ANOVA with degrees of freedom = 1,19; p = achieved p-value for reported effect from ANOVA.

Neuroimaging

Categorization Regions-of-Interest. Table 6 lists the voxel coordinates (Talairach & Tournoux, 1988) for the maximum intensity voxel of the 17 ROIs in which significant changes were observed for the categorization tasks relative to the control tasks. These coordinates are also the center of mass for the spherical ROIs created and used for further statistical analysis of the functional images.

After identification of the ROIs, the eight *net* activation values (DEL, DFL, PEL, PFL, DET, DFT, PET, PFT) for the 17 ROIs were further investigated via ANOVA. The results of the ANOVA for regions with statistically significant effects are reported in the last 3 columns of

Table 6. As reported in Table 6, seven regions were found to have a significant interaction of task phase and training (LTxFX), 2 regions were found to have a significant main effect of training (feedback versus exposure), and 1 region, the left middle temporal gyrus, was found to have significant main effects of both task phase (learning versus testing) and training (feedback versus exposure), but not a significant interaction of the two factors. None of the regions were found to have a significant main effect of category assignment (deterministic versus probabilistic) in the absence of a significant interaction.

One goal of Experiment 2 was to isolate differences in the category learning condition (feedback versus exposure), category assignment (deterministic versus probabilistic) and, category task phase (learning versus testing) – three factors that had either been confounded within previous experiments or were often not explicitly accounted for when making comparisons across studies. One conclusion that can be drawn from Experiment 2 is that whereas differences in the learning condition and category task phase are found to be associated with significant differential neural activation, differences in category assignment (deterministic versus probabilistic) – at least with respect to this particular categorization task – are not. The primary result reported for the ROIs investigated in this experiment is that there is a significant interaction of category learning condition and category task phase since an interaction of these two experimental factors was reported for 7 of the 17 ROIs.

An example of this interaction is illustrated in Figure 6. In this ROI, activation in the left

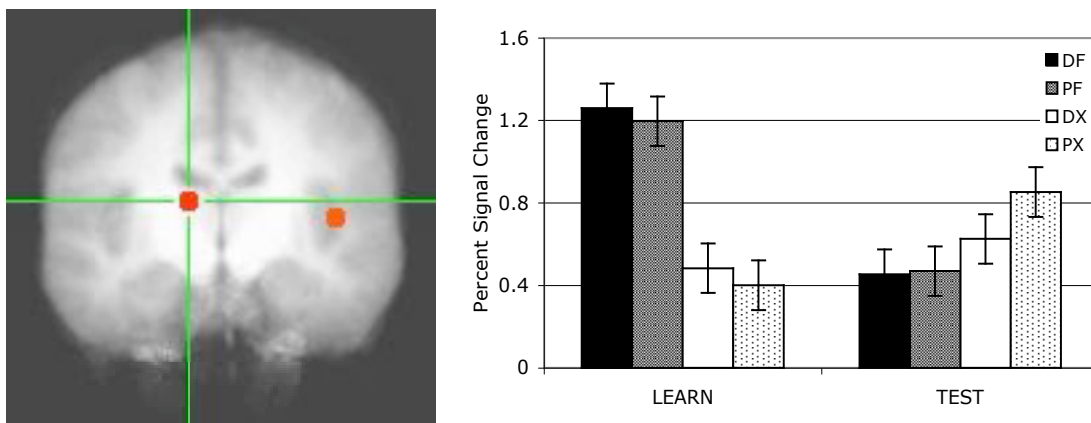


Figure 6. LTxFX interaction - left thalamus ROI. The panel on the left shows the location of the ROI in the coronal plane. The right panel reports the percent signal change for each experimental condition (net of control). Structural image is an average image from all subjects. Error bars are rmse.

thalamus was greater during feedback learning (black and gray bars in the LEARN condition) relative to exposure learning regardless of the category structure. A similar pattern of activation was found in all of the ROIs reporting a significant LTxFX interaction. Four of the ROIs reporting this interaction were associated with the fronto-striatal learning system and this pattern of activation would be generally consistent with the proposed role of the striatal system in feedback learning tasks.

The LTxFX interaction was also observed in three occipital regions including two left lingual gyrus ROIs and the left cuneus (all BA 18). Since the visual characteristics of the stimuli in Experiment 2 were the same for all conditions, differences in neural activation in these visual perception regions cannot be attributed to perceptual differences in the stimuli as in Experiment 1. Reber and colleagues have previously reported differential activation of occipital cortex associated with processing of categorical stimuli (Reber et al., 1998a, 1998b; Reber et al., 2002). In these studies, areas of relative deactivation in occipital cortex were reported for a learned category relative to both a contrast category and random, noncategorical, items. However, in Reber's paradigms, subjects were exposed to the categorical stimuli multiple times prior to testing and were not exposed to the contrast category prior to testing. Therefore, the deactivation of occipital cortex reported for the learned category relative to the contrast category stimuli may have also been due the relative amount of exposure subjects had received to the stimuli prior to testing. Relative to the same baseline, previously seen items would show reduced activation due to priming or habituation (Buckner et al., 1998) and new, noncategorical stimuli would show greater activation as a result of their novelty (Gabrieli, Brewer, Desmond, & Glover, 1997) and therefore, a direct comparison of activation for the two stimulus conditions could result in a deactivation for the category relative to the contrast category conditions. In the current study, exposure to the experimental and control stimuli was equated, eliminating priming and novelty effects as possible explanations for differences in activation between the various experimental conditions. One conclusion that can be drawn from the current study with regard to these occipital activations is that neural regions normally associated with low-level perceptual processing can show differential activation with respect to higher-level cognitive task demands, possibly through the action of attentional modulation (Vidnyanszky & Sohn, 2005) and/or recurrent processing loops (Lamme, 2000, 2002; Lamme & Roelfsema, 2000).

As previously noted, the conclusions that could be drawn from the results of Experiment

1 were limited since more than one experimental factor varied between the task conditions. Seven ROIs in Experiment 1 showed a significant main effect associated with activation during the prototype-distortion task relative to the probabilistic-cue task. Since both tasks in Experiment 1 were learned using a feedback paradigm, the difference in activation should be related to one of the factors that differed between the two TASK conditions such as the category structure (deterministic versus probabilistic) or the stimuli (polygons versus cards). The results of Experiment 2 appear to eliminate differences in category structure as being the source of the differential activation in Experiment 1 since in Experiment 2 no significant main effects of category structure were found.

One possible concern with the paradigm used in Experiment 2 is that subjects were tested using only a single category feedback structure within a session (one-half were tested using probabilistic category structures in the first session and one-half were tested using deterministic category structures in the first session). This approach was taken to reduce the possibility of subject confusion with respect to the various learning conditions. However, the results would indicate that this blocking of category structure within session had no effect. In general, activation for probabilistic and deterministic conditions (which were manipulated between sessions) was similar across conditions whereas activation for exposure versus feedback learning conditions (which were manipulated within each session) was associated with differences in activation. If blocking the fMRI sessions by category structure had an effect on subject performance, I would have expected main effects associated with deterministic versus probabilistic learning to be exaggerated.

A second limitation of this study is that categorization was observed only during the early stages of learning rather than during both early and late epochs of categorization as in Experiment 1. However, the decision to limit Experiment 2 to the early stages of learning was motivated by the extant neuropsychological literature and fMRI experiments in which differences were identified in early, but not late, performance. Since the fMRI results of Experiment 1 failed to show a main effect of EPOCH, limiting Experiment 2 to the early epochs of category learning and testing allowed me to accommodate three experimental conditions. It is still possible that if training was extended for longer periods of time differences in activation between conditions related to category acquisition over time might be observed. This question will be further addressed within the context of a single categorization task in Chapter IV.

Finally, it should be noted that, although there was no difference associated with category assignment in Experiment 2 (deterministic versus probabilistic), this null result may be a function of the particular stimuli that were used. One possibility may be that this particular type of stimuli, one-to-three cards, lends itself to a rule-based strategy but that deterministic and probabilistic category assignment would result in differential activation if a different type of stimuli, such as single objects with binary features, were used. For instance, Hopkins and colleagues (Hopkins, Myers, Shohamy, Grossman, & Gluck, 2004) have tested amnesiac (hypoxic) subjects using both card stimuli and a set of object stimuli with multiple features. In both cases, category feedback was assigned using probabilistic-cues. However, in this study, the amnesiac subjects were impaired during early training on a probabilistic-cue task when the integral object stimuli were used but, consistent with prior studies, the amnesiac subjects performed comparable to controls during early training when the card (non-integral) stimuli were used providing evidence that the type of stimuli, may have an impact on how the task is performed.

The results from Experiment 2 provide additional insight into the neural correlates of category learning and testing. Whereas I did find evidence of neural regions that appear to be differentially sensitive to changes in learning condition and learning phase, contrary to the conclusions drawn from previous studies, I did not find evidence that, in the case of card stimuli, neural regions are differentially sensitive to differences in category structure.

CHAPTER IV

CHANGES IN NEURAL CORRELATES OVER THE COURSE OF CATEGORY ACQUISITION

Experiment 3

Taken together, the results of Experiments 1 and 2 generally support the conclusion that differences in neural activation are associated with differences in explicit categorization tasks. However, based on behavioral, modeling and behavioral neuropsychological studies, I also expected to find evidence of differential neural activity with respect to early versus late category acquisition. Contrary to this expectation, changes in neural activity associated with early versus late category acquisition were not found to be statistically significant in Experiment 1 and this condition was not tested in Experiment 2.

Experiment 3 revisits the issue of changes in categorization over the course of category acquisition by examining changes in activity within a single categorization paradigm (Experiment 1 examined two paradigms) utilizing a within-subjects design (Experiment 1 examined this factor between subjects) and a 3-Tesla scanner (Data for Experiment 1 were acquired with a 1.5-Tesla scanner) to acquire the data. As discussed in Chapter II, shifts in categorization behavior have been viewed as evidence of various hybrid categorization systems in which one type of category system or representation is used during early epochs of the task and another system or representation is used during later epochs of the task. Whereas Experiments 1 and 2 demonstrated that neural activity may change with respect to differences in the categorization task, the goal of Experiment 3 is to examine how neural activity may change over the course of category acquisition.

To maximize my ability to identify a change in neural activation over time, I utilized a behavioral paradigm in which change in category behavior over time has been documented in several behavioral studies. Evidence from these empirical and modeling studies has supported a hybrid model of categorization in which a rule-based categorization system accounts for judgements made during early epochs of categorization (Johansen & Palmeri, 2002; Nosofsky & Palmeri, 1998) and an exemplar-based system for late epochs of the same task (Johansen & Palmeri, 2002). This hybrid model, as the name implies, evolved by combining aspects of single-

system models of categorization. As noted in Chapter II, hybrid models have been able to address patterns of behavioral data that are not well-described by single-system models by positing multiple-systems that utilize different processes or representations for each system (Ericksen & Kruschke, 1998; Johansen & Palmeri, 2002; Johnstone & Shanks, 2001; Nosofsky, Palmeri et al., 1994; D. J. Smith & Minda, 1998). Using hybrid models, researchers have been able to explain categorization phenomena that include a bias toward the use of verbal rules in categorization (Ashby et al., 1998), distributions of individual-subject generalization patterns for ill-defined categories (Johansen & Palmeri, 2002), and performance on categorization of prototypes versus exceptions (D. J. Smith & Minda, 1998; D. J. Smith, Murray, & Minda, 1997).

The behavioral paradigm adapted for use in the current experiment actually arose from a series of studies investigating the use of rule-based decisions in categorization. In response to a pattern of behavioral categorization data that was not well specified by a single-system model, Nosofsky, Palmeri and McKinley (1994) proposed a rule-plus-exception (RULEX) model of classification learning. Classification decisions in the RULEX model were made by having the model form a rule and then remember (store) exceptions to that rule. Subject to a criterion, the model would search for perfect single-dimension rules, imperfect single-dimension rules and conjunctive rules to classify stimuli. If successful at any of these rule-testing stages, the model could continue with the successful rule and then learn exceptions to that rule (Nosofsky, Palmeri et al., 1994).

One set of behavioral data to which the RULEX model was fit was a set of results reported by Medin and Schaffer (1978) that had previously been used to support the context model: a single-system exemplar model of categorization. Subjects in the Medin and Schaffer study were tested on stimuli that had four binary-valued features. The category structures used to assign stimuli to the categories in this series of experiments were ill-defined in that a simple rule could not be formed to correctly categorize the items into two categories. Both the context model (Medin & Schaffer, 1978) and RULEX (Nosofsky, Palmeri et al., 1994) provided good fits to the average classification data reported by Medin and Schaffer. However, Nosofsky et al. extended their analysis to examine the generalization patterns for individual subjects. A generalization pattern is the pattern of responses made to new, transfer stimuli, after receiving feedback training and these patterns can be indicative of the manner in which a subject is making categorization decisions. Nosofsky et al. (1994) tested whether RULEX could predict a distribution of

generalization patterns by collecting data on a large number of subjects using the same category structures used by Medin and Schaffer (1978). RULEX and the context model were fit to the distribution of observed generalization patterns after 16 blocks of training and RULEX was able to account for 85.7% of the variance compared to only 35.9% accounted for by the context model (Nosofsky, Palmeri et al., 1994). Thus, while both RULEX and the context model were able to account for the average transfer data in categorization paradigms, RULEX was also able to account for individual subject variability that was masked by averaging the data.

In the RULEX model only rules and exceptions were stored for use in making category decisions. However, the RULEX model would then predict no recognition memory for items unless they were exceptions to the rule – a result inconsistent with known human behavior. Palmeri and Nosofsky (1995) therefore extended the RULEX model to account for recognition data by allowing for the storage of some residual memory for non-exception items and allowing special weighting of exception items in the summed similarity calculation for making recognition decisions. While the generalization patterns still showed a noticeable proportion of the subjects using imperfect single-dimension rules to categorize transfer items, the generalization pattern after 25 blocks of training also indicated that the largest proportion of subjects used an exemplar-based strategy. Nosofsky, Palmeri and McKinley (1994) had also reported the same exemplar-based pattern in their generalization pattern after 16 blocks of training but in their study the proportion of subjects observed to be using the exemplar-based strategy had been less than either of the rule-based conditions. The emergence of these exemplar-consistent generalization patterns in the later blocks of category learning is consistent with theories that posit the creation of memory traces with each exposure to an instance of an item. Under these theories, exposure to items can eventually lead to automatic memory-based responses to that item (Logan, 1988, 1992, 2002; Palmeri, 1997).

Johansen and Palmeri (2002) sought to explore the basis of the exemplar generalizations reported by Nosofsky and colleagues (Nosofsky, Palmeri et al., 1994; Palmeri & Nosofsky, 1995) by modifying these earlier paradigms in two ways. First, they extended the number of training blocks to 32. Second, they introduced testing of transfer blocks at intervals during the training (rather than only at the end) in order to determine how the rule-based and exemplar generalizations were changing over time. In this study, subjects learned to categorize nine stimuli as either members of category A or category B in a standard feedback training paradigm. Each

pass through the nine stimuli comprised one training block. After subjects had completed a certain number of training blocks, they were asked to categorize the 9 training stimuli as well as 7 additional transfer stimuli without receiving feedback. The subject's pattern of A/B category judgments for the 7 transfer stimuli (referred to as their generalization pattern) was interpreted as a behavioral marker indicating whether the subject was making decisions consistent with a rule-based or exemplar-based model of categorization. Using this modified paradigm, Johansen and Palmeri (2002) reported behavioral and theoretical modeling evidence that they interpreted as supporting a shift in representation from single diagnostic dimensions, or simple rules, during early learning, to multiple-dimensions, or exemplars, during late learning.

Johansen and Palmeri (2002) interpreted their results in terms of a representational shift from a rule-based representation during early categorization epochs to an exemplar-based representation during late epochs. However, the rules-to-exemplars interpretation was not the only possible explanation. For instance, the behavioral data could also reflect a categorization process in which selective attention to dimensions of the stimulus changed over time as suggested by Nosofsky and Johansen (2000) who have been able to fit an exemplar model to the generalization pattern reported for Experiment 2 of Palmeri and Nosofsky (1995). Thus, hybrid models of categorization have developed to explain more complex patterns of behavioral data but questions remain with respect to the nature of the processes underlying these models. Since one of the acknowledged limitations of mathematical models is that different models can often predict the same behavior (Olsson et al., 2004; Pitt & Myung, 2002), Experiment 3 was designed to examine the neural processes associated with this shift in categorization behavior over the course of category acquisition by searching for affirmative evidence of a rules-to-exemplars process dissociation (Poldrack, 2000).

Functional neuroimaging studies have previously associated rule use in categorization, hypothesis testing, and guessing with specific neural regions in the frontal and parietal lobes as well as the cingulate and striatal system (Elliott & Dolan, 1998; Elliott et al., 1999; Patalano et al., 2001) and long-term memory primarily with the medial temporal lobe (Scoville & Milner, 1957; Squire & Knowlton, 2000). Experiment 3 will focus on examining these neural regions for evidence of a shift in functional activation over the course of category acquisition. As noted above, evidence of a shift in representation from rules-to-exemplars will be sought in the form of a process dissociation in the neuroimaging results of normal subjects (Poldrack, 2000). In FMRI,

a process dissociation occurs when the activation in two neural regions changes in opposite directions relative to some independent variable, such as training epoch. For instance, if a rule-based system is used during early, but not late, epochs of categorization, I should find differential activation for categorization relative to control during early - but not late - epochs of the task in neural regions previously associated with cognitive processes such as rule application, hypothesis testing, and/or guessing. The pattern of results expected for MTL activations under exemplar theory is slightly more complicated in that there will be activation associated with both exemplar storage and exemplar retrieval. If distinct neural regions are differentially active with encoding and retrieval, then I might expect two patterns of activation within the MTL, a pattern of decreasing activation over time associated with exemplar encoding and a second pattern of increasing activation over time associated with exemplar retrieval. Therefore, in addition to an analysis similar to the one used in Experiments 1 and 2, I will also use an ROI analysis to test for patterns of activity consistent with a rules-to-exemplars process dissociation.

To test for this process dissociation, I chose to replicate Johansen and Palmeri's (2002) original task as closely as possible while still meeting the methodological demands of a functional imaging paradigm. In adapting the Johansen and Palmeri paradigm, I made three modifications. First, since subjects would generally make more errors during early learning relative to late learning, they would receive different proportions of corrective feedback during different learning epochs. To control for this possible confound, I included two types of categorization trials, learning and testing trials. Subjects received corrective feedback during learning trials but did not receive corrective feedback during testing trials. A second adjustment made to the original behavioral paradigm was to yoke presentation of control stimuli to presentation of the learning and testing stimuli, thus equating the control, learning, and testing conditions for number of exposures and eliminating any effects related to priming (Buckner et al., 1998; Newman, Twieg, & Carpenter, 2001). A final adjustment to the paradigm was to use a fixed timing sequence for the learning, testing, and control trials rather than allow subjects an unlimited amount of time in which to make a categorization decision and response. This change from the original paradigm allowed us to equate total exposure time for the stimuli within and between subjects. Pilot subjects were tested to assure that they were able to acquire the category using these fixed presentation and response time limits.

Method

Subjects

Subjects were 11 individuals affiliated with Vanderbilt University who received cash payment for their participation. Each subject participated in two separate fMRI sessions approximately 6 weeks apart. All subjects were tested individually and self-reported to be right-handed with normal or corrected-to-normal vision. Written informed consent was acquired from each individual before each testing session.

Apparatus and Image Acquisition Parameters

Imaging was performed on a GE 3T SIGNA research MRI scanner fitted with a high performance local head gradient and RF coils. Functional T2*-weighted images were acquired using a gradient echo, single-shot echo planar imaging sequence (GR-EPI) with a matrix size of 64 x 64 (FOV 240mm), echo time of 60 ms, flip angle of 70 degrees, TR of 2000 ms, in-plane resolution of 3.75 x 3.75, and slice thickness of 7 mm with 0 mm gap. 18 axial slices were placed starting at the ventral surface of the temporal lobe and extending in the superior plane to provide whole brain coverage. Each subject completed between four and eight functional imaging scans during which both behavioral and imaging data were collected. For anatomical localization, a standard whole-brain, T1-weighted, 3D FSPGR sequence was acquired after functional data collection (184 sagittal slices, in-plane resolution of 0.9375, slice thickness of 1mm, no gap).

Stimuli

Stimuli were four sets of objects (rockets, lamps, sailboats, and neckties – see Appendix B for details regarding stimulus construction) randomly paired and assigned to the two conditions for each subject.

Within each condition, one set of objects was assigned to the category learning (LEARN) and testing (TEST) conditions and the other was assigned to the control condition (CNTRL). Each set of objects varied along four binary-valued dimensions (Medin & Schaffer, 1978). Assignment of physical dimensions and features to the abstract category structure for each object group was randomized for each subject.

Stimuli were assigned to Category A (5 items) and Category B (4 items) as illustrated in the top and bottom panels of Table 7. Stimulus presentation was via goggles worn within the scanner headcoil. Resolution of the goggles was set to 1024 x 768. At this resolution, the stimuli subtended approximately 5 degrees of visual angle.

Table 7
Category Structures used in Experiment 3

Medin & Schaffer (1978) Exp 2,3; Johansen & Palmeri (2003) Exp1			
	Category A		Category B
A1	0 0 0 1	B1	0 0 1 1
A2	0 1 0 1	B2	1 0 0 1
A3	0 1 0 0	B3	1 1 1 0
A4	0 0 1 0	B4	1 1 1 1
A5	1 0 0 0		

Johansen & Palmeri (2003) Experiment 3			
	Category A		Category B
A1	0 0 0 1	B1	0 0 1 1
A2	0 1 0 1	B2	1 1 0 1
A3	0 0 0 0	B3	1 1 1 0
A4	0 1 1 0	B4	1 0 1 0
A5	1 0 0 0		

Each object has four binary-valued dimensions. 0 represents one value on the dimension and 1 represents the other value. A1 through A5 were assigned to the A category; B1 through B4 were assigned to the B category. The 7 remaining stimuli (16 total possible stimuli for each stimulus set minus the 5 stimuli assigned to category A and the 4 stimuli assigned to category B) were not used.

Procedure

All subjects were tested in both conditions of the experiment in two separate functional imaging sessions completed approximately six weeks apart. Due to a scanner malfunction, data from the second scan session for one subject was lost. The only difference between the two scan sessions was the underlying category structure of the stimuli. In the first scan session I used the category structure from Experiment 3 of Johansen & Palmeri (2002) (bottom panel of Table 7) whereas in the second scan session I used the category structure from Experiment 3 of Medin and Schaffer (1978) (top panel of Table 7).

Pre-Scan Familiarization Session. An hour before the scan session, subjects completed a short, pre-scan familiarization session. During this session, task instructions were presented via a short, self-paced, computer presentation. As part of this presentation, subjects practiced making categorization judgments in both untimed and timed categorization trials using stimuli (chairs and abstract designs) that were never used for the experimental trials.

Functional Imaging. LEARN, TEST, and CNTRL stimuli were always presented in blocks of 9 trials. Each of the 9 categorical stimuli were presented during each block with the order of stimulus presentation for each block based on a random sample of the 9 possible stimuli

EQ	F	LR	C	TS	C	F	LR	C	TS	C	F	C	TS	C	LR	F	C	TS	C	LR	F
----	---	----	---	----	---	---	----	---	----	---	---	---	----	---	----	---	---	----	---	----	---

Figure 7. Schematic of block design used in Experiment 3. EQ: Equilization – 20 seconds of fixation at beginning of run, F: Fixation – 16 seconds of fixation baseline, LR: block of LEARN trials, TS: block of TEST trials, C: block of CNTRL trials. The order of experimental blocks in the first half of the fMRI scan is mirrored in the second half of the scan.

drawn without replacement. The CNTRL stimuli were matched with the LEARN and TEST stimuli for number and length of exposure and provided a non-categorization baseline for purposes of identifying ROIs. The entire scan session included 8 functional MRI scans during which a total of 32 blocks of each stimulus type were presented. As illustrated in Figure 7, each separate functional imaging scan included 4 blocks of LEARN trials during which subjects received corrective feedback, four blocks of TEST trials during which subjects received no corrective feedback and 8 blocks of CNTRL trials presented in an ABBA design to control for scanner drift (Pettersson et al., 1999; Skudlarski et al., 1999).

LEARN Trials. Following Medin and Schaffer (1978) and Johansen and Palmeri (2002), subjects learned the category during the LEARN trials in a standard category learning procedure. For each trial, the subject was presented with one of the 9 category stimuli (5 A's, 4 B's). Each trial started with a 250 ms fixation cross after which the stimulus was presented with the question 'A or B ?' immediately below the stimulus. Subjects had up to 1750 ms to respond before corrective feedback was presented. If the subject responded prior to the 1750 ms deadline, the stimulus remained on the screen and the corrective feedback was immediately displayed beneath the stimulus. If the subject made a correct categorization decision, the corrective feedback would state 'Correct'. If the subject made an incorrect categorization decision, the corrective feedback would state 'Incorrect it's a ___' with the name of the correct category. Responses were made via a button response box strapped to the subject's right hand. Subjects were instructed to press their first finger for a category 'A' response and to press their middle finger for a category 'B' response. If the subject did not respond within the 1750 ms time limit, the feedback screen stated 'no response detected, it's a ___' with the name of the correct category response. Subjects were told during the familiarization session that if they received feedback that no response was detected they either did not respond within the response deadline limit or their fingers had slipped off of the appropriate response box keys and that they should make appropriate corrections on future trials.

TEST Trials. The second type of trials that subjects completed were TEST trials. TEST trials were the same as LEARN trials (same stimuli, same time constraints) except that subjects did not receive corrective feedback. Instead, subjects would see either ‘response collected’ if they had responded within the 1750 ms time limit or ‘response not detected’ if the response was not made within the deadline.

CNTRL Trials. The presentation of control stimuli was yoked to the categorization stimuli (LEARN and TEST) in order to provide a second baseline condition that was equated to these categorization conditions for visual complexity and number of exposures. Therefore, for each block of nine LEARN or nine TEST trials the subject would also complete a block of nine CNTRL trials. CNTRL trials differed from the categorization trials in two ways. First, the control stimuli came from a different set of objects than the LEARN and TEST stimuli. Second, subjects were asked to ‘Press a Key’ rather than make a categorical ‘A or B?’ judgement. Like the TEST trials the subject's response was simply acknowledged as ‘response collected’ or ‘response not detected’.

Analysis

Functional Imaging Analysis

All functional imaging data were analyzed using the AFNI (Analysis of Functional NeuroImages) software package (Cox, 1996).

Individual Subject Data. Prior to statistical analysis of the functional images, each subject’s time-dependent functional imaging data were examined and corrected for large between phase fluctuations due to scanner artifacts and subject movement. Images from each functional run were then registered to a single base image to correct for in-plane and out-of-plane subject movement. The six motion parameters used to correct each phase of the imaging run were recorded and used as covariates in the statistical analysis to determine percent signal change for each condition. Each subject’s structural image was also registered to the same functional base image.

Each individual's motion-corrected functional imaging runs were then reduced to a measure of percent signal change relative to the fixation baseline for each condition. To convert activation to a measure of activity associated with each task a stimulus function of 1's and 0's representing condition and non-condition time points was created for each of the task conditions

in each run (LEARN, TEST, and CNTRL) and then convolved with a canonical hemodynamic response function. The convolved condition functions and the six motion parameters obtained from the image registration process were then submitted as independent variables (IVs) with the voxel activation value as the dependent variable (DV) to a multiple regression analysis. The full model included an estimate of baseline in addition to the four IV's. The output of this analysis for each individual was then converted to a measure of percent signal change and placed into the coordinate system of Talairach and Tournoux (Talairach & Tournoux, 1988).

Group Analysis. Prior to creating ROI maps or completing statistical analyses, an analysis was completed to determine whether I could collapse conditions across the two scanning sessions. The only difference in the experimental conditions between the scan sessions was that different category structures were used (see Table 7). A comparison of functional activation for the LEARN, TEST, and CNTRL conditions in sessions 1 versus session 2 revealed no statistical differences within conditions thereby allowing us to collapse data across the two category structures increasing the statistical power of the tests to find differences associated with the remaining independent variables.

ROI Analyses. I performed two sets of analyses, one based on functionally determined ROIs and a second analysis based on *a priori* ROIs.

Maximum Intensity ROI Map. To provide continuity between the analyses performed on the data in Experiments 1 and 2 and the current experiment, I created a initial functional map to identify ROIs based on Maximum Intensity for categorization relative to the control conditions. To identify these ROIs, the percent signal change values for each subject in both TASK conditions (LEARN and TEST) and the CNTRL condition were submitted to an ANOVA to determine voxels in which there was a significant main effect (corrected for multiple comparisons through a combination of statistical threshold and cluster size) of categorization relative to control. Regions of maximum intensity within this map were then identified and spherical ROIs (257 vmul) were created at these voxel coordinates for use in additional statistical analysis of the data.

Rule-based ROI Map. To identify areas specifically associated with rule-type activity, an ROI map was created by placing spherical ROIs with a 4mm radius (Patalano et al., 2001; Poldrack et al., 2001) at the Talairach coordinates reported in studies of rule application (Patalano et al., 2001), hypothesis testing (Elliott & Dolan, 1998), and guessing (Elliott & Dolan,

1998, 1999). These ROIs were then investigated for evidence of differential activity with respect to LT and RUN.

MTL ROI Map. The procedures used to identify areas specifically associated with MTL activity required several steps. The first step in generating the MTL ROIs was to create two separate MTL ROI maps to identify areas of differential activity with respect to either early (Early ROI map) or late (Late ROI map) categorization versus control. The Early map was the result of a paired *t-test* comparing the LEARN and TEST conditions of runs 1 and 2 with the CNTRL condition of runs 1 and 2. The Late map was the result of a paired *t-test* comparing the LEARN and TEST conditions of runs 7 and 8 with CNTRL conditions of runs 7 and 8. To correct for multiple comparisons, only voxels reaching a statistical significance threshold of $p < .0001$ and in contiguous clusters exceeding 300 vmul were included in the final Early and Late ROI maps resulting in areas with statistical significance of $p < .05$ corrected for multiple comparisons.

The next step after creation of the Early and Late ROI maps was to identify areas of conjunction between the two maps. Therefore, for each voxel in the MTL, there were four possibilities with respect to the Early and Late functional ROI maps. A voxel could be included 1) only in the Early map, 2) only in the Late map, 3) in both the Early and Late maps or 4) in neither the Early or Late map. The thresholded Early and Late ROI maps were therefore combined and further analyzed to identify areas associated 1) *only* with early categorization minus control resulting in the EO ROI map 2) *only* with late categorization minus control resulting in the LO ROI map and 3) *both* early and late categorization minus control resulting in the EandL ROI map. It should be noted that the process of creating conjunction maps could result in clusters with volumes of less than 300 vmul. However, since these smaller clusters were originally included in maps thresholded at $p < .05$, corrected for multiple comparisons, they have been retained for further statistical analysis.

Results & Discussion

Behavioral

Behavioral data (percent correct and reaction time) were collected concurrent with collection of the functional images. As illustrated in Figure 8, subjects were able to learn the category across the series of eight training runs. ANOVA confirmed a significant effect of RUN

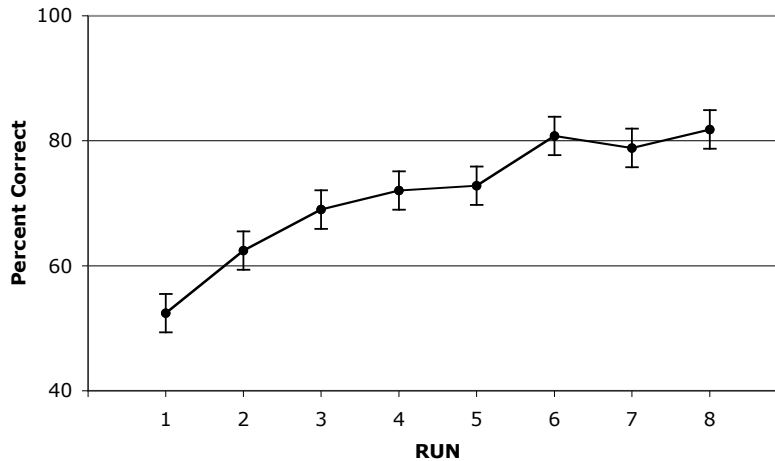


Figure 8. Mean percent correct for categorization performance in LEARN and TEST conditions for each run (1 through 8) of Experiment 3. Performance across runs improved as confirmed by ANOVA [$F(7,68) = 27.63, p < .05$]. Error bars are *rmse*.

(1 through 8) [$F(7, 142) = 27.633, p < .05$], no effect of category structure (J&P vs M&S) [$F(1,142) = 1.60, p > .05$], and no significant interaction of RUN and category structure [$F(7,142) = 0.444, p > .05$]. Since the ANOVA revealed no statistical differences between the J&P and M&S Figure 8 reports the combined mean percent correct for all subjects for both Session 1 and Session 2 data collection. An ANOVA performed on mean RTs also confirmed a significant main effect of RUN (1 through 8) [$F(7, 142) = 2.257, p < .05$].

Since the accuracy data indicated no statistically significant difference in performance between the J&P and M&S conditions I also tested whether there was a functional difference between the two category structures. Consistent with the behavioral results, paired *t-tests* for each condition confirmed there were no statistically significant differences in neural activation between sessions using the J&P or the M&S category structures and therefore, for purposes of the neuroimaging analysis, the data from these two conditions were treated as one condition.

Maximum Intensity ROI Analysis

Consistent with the analysis of Experiments 1 and 2, I first identified ROIs corrected for multiple comparisons based on the comparison of the categorization conditions (LEARN and TEST) relative to the control condition (CNTRL). The results of this ROI analysis are presented

Table 8
Experiment 3 Regions of Interest and results of ANOVA

LR	AP	IS	Hemi	Brain Region	BA	Effect	F	p
<u>Frontal</u>								
50	15	26	Right	Middle Frontal Gyrus	9	LT	3.957	.048
35	21	-5	Right	Inferior Frontal Gyrus	47			
44	29	24	Right	Middle Frontal Gyrus	46			
-38	5	29	Left	Inferior Frontal Gyrus	9			
-41	24	18	Left	Middle Frontal Gyrus	46	RUN	2.873	.006
-13	45	38	Left	Superior Frontal Gyrus	8			
<u>Limbic</u>								
-2	-40	34	Left	Cingulate Gyrus	31			
2	38	0	Right	Anterior Cingulate Gyrus	24			
0	15	39	Left	Cingulate Gyrus	32			
11	-5	14	Right	Thalamus	VAN	LT	15.604	.000
-12	-7	6	Left	Thalamus	VLN	LT	15.282	.000
-6	-24	-4	Left	Red Nucleus		LT	22.575	.000
<u>Temporal</u>								
-51	-59	31	Left	Superior Temporal Gyrus	39			
-33	-57	-17	Left	Fusiform Gyrus	37	LT	17.323	.000
-41	-60	-4	Left	Middle Temporal Gyrus	37			
-30	23	-2	Left	Insula	13			
35	-17	18	Right	Insula	13	LT	8.776	.003
<u>Parietal</u>								
5	-64	42	Right	Precuneus	7			
-24	-63	38	Left	Precuneus	7			
<u>Occipital</u>								
5	-77	-14	Right	Lingual Gyrus	18	LT	17.524	.000
						RUN	2.254	.030

Regions of Interest based on maximum F values for categorization conditions relative to control conditions. R = Right (+), L = Left (-); A = Anterior(+), P = Posterior(-), S = Superior (+), I = Inferior (-); Hem = Hemisphere; Brain Regions = Talairach Label; BA = Brodmann Area; * no Brodmann label. Effects = ROIs with significant effects confirmed by ANOVA; LT = Main Effect of Learn vs Test; RUN = Main Effect of RUN 1 through RUN 8; F = F value for reported effect from ANOVA with degrees of freedom = x,x (LT) or x,x (RUN); p = achieved p-value for reported effect from ANOVA.

in Table 8. A main effect of TASK was found for 7 of the 20 regions. The difference isolated between these two conditions was whether subjects were receiving feedback (LEARN) or not (TEST) after they made their categorization judgement. As in Experiments 1 and 2, there is differential activation with this manipulation in the thalamus and lingual gyrus. Using these ROIs, a main effect of RUN was identified only in the left middle frontal gyrus (BA 46), an area that has been associated with on-line monitoring and manipulation of cognitive representations (Petrides, 1996, 2002) and the right lingual gyrus (BA 18). I had reported two regions of the

leftlingual gyrus to be differentially active with the feedback relative to no feedback in Experiment 2 and suggested that this activation may be related to an attentional effect related to stimulus features (Vidnyanszky & Sohn, 2005) and/or recurrent processing of stimuli via feedforward and feedback circuits (Lamme, 2000; Lamme & Roelfsema, 2000). Taken together, the results of this overall analysis are consistent with Experiment 1 and 2 and neither support nor refute the rules-to-exemplars hypothesis. Therefore, I will turn to examination of regions based on *a priori* ROIs for additional evidence of a process dissociation.

A Priori Rule and MTL ROIs

As noted in the introduction of this experiment, if categorization is associated with a shift from a rule-based system to an exemplar (or other) system, one pattern of results that would provide strong evidence of this shift would be a process dissociation in which areas of frontal cortex previously associated with rules would show decreasing activation and areas of MTL previously associated with exemplar retrieval would show increasing activation as experience with the stimuli increases.

To test for significant activation in areas previously associated with rules, I created ROIs from coordinates reported in studies of rule application in categorization (Patalano et al., 2001), hypothesis testing (Elliott & Dolan, 1998), and guessing (Elliott et al., 1999). I then interrogated these regions using an ANOVA to test for main effects of TASK (Learn versus Test) and RUN (1 through 8) as well as the interaction of TASKxRUN. In addition, for regions that showed a significant main effect of RUN, I also tested whether the linear contrast was significant. Summary results of this analysis are reported in Table 9.

A rules-to-exemplars system could operate in one of several ways. In one scenario, the rule-based system may be automatic and compete with the exemplar system to produce a response. In this case, I could expect activity associated with rule application to remain fairly constant over the entire period of category acquisition since the system would continue to generate and apply the rule to supply the rule-based response regardless of whether it ultimately won or not. In this case, it would be relatively difficult to make any strong conclusions, based on neural activation, as to whether categorization was being supported by the rule-based or exemplar-based system. I will refer to this pattern of activation as the "weak" version of the rules-to-exemplars hypothesis. However, a second possibility is that the rule-based system could

Table 9
Rule ROIs with Significant Effects

Study	LR	AP	IS	Hem	Brain Region	BA	Effect	F	p
G	-37	35	24	L	Middle Frontal Gyrus	10	RUN	2.246	.031
G	15	26	-12	R	Inferior Frontal Gyrus	47	RUN	2.757	.009
H	-8	22	48	L	Medial Frontal Gyrus	8	RUN	2.426	.020
R	39	19	25	R	Middle Frontal Gyrus	9	RUN	2.704	.010
G	6	33	24	R	Anterior Cingulate	32	LT	6.257	.013
H	46	6	16	R	Insula	13	LT	4.527	.034
P	-33	-5	46	L	Middle Frontal Gyrus	6	LT	5.184	.024
R	1	-13	0	R	Thalamus	*	LT	29.86	.000
H	10	-13	7	R	Thalamus	*	LT	10.82	.001
G	-45	-51	45	L	Inferior Parietal Lobule	40	LT	4.762	.030

Regions significantly active for categorization conditions relative to control conditions. Study = study from which the ROI used in the conjunction analysis was identified. P = (Patalano et al., 2001); H = (Elliott & Dolan, 1998); G = (Elliott et al., 1999); R = Right (+), L = Left (-); A = Anterior (+), P = Posterior (-); S = Superior (+), I = Inferior (-); Hem = Hemisphere; Brain Region = Talairach Area; BA = Brodmann Area.; * = no Brodmann Area; Effect = Factor with Significant Effect; F = value of F-statistic; p = p-value associated with F statistic. RUN = Main effect of RUN; LT = Main effect of TASK (Learn versus Test).

operate during early acquisition with activation decreasing as a second system acquired enough information to generate a response based on memory traces (exemplar) or an abstracted representation of previously seen items (prototype). This pattern of responses would provide stronger evidence of a shift in categorization from an early rule-based system to a late exemplar-based system. I will refer to this pattern of activation as the "strong" version of the rules-to-exemplars hypothesis. The four regions reported in Table 9 with a significant main effect of RUN also had a significant linear trend ($p < .05$) with greater activation reported during early than late epochs of the task. This is the pattern of results predicted for rule-based ROIs in the strong version of a rules-to-exemplars process dissociation.

At this point, I have identified areas with activation patterns consistent with one-half of the strong version of the rules-to-exemplars process dissociation – that is, I have identified four neural regions previously associated with rule-use, hypothesis testing, or guessing that are associated with significant linear changes (decreases) in activation with increased experience with the stimuli. I now turn my attention to evidence of exemplar encoding and retrieval during the categorization task.

Unlike the rule retrieval condition, I do not have a limited set of studies that would result in a finite set of exemplar ROIs. However, since exemplars are presumed to be encoded and retrieved from LTM, exemplar encoding and retrieval should be associated with activity in the MTL. I tested this hypothesis by looking for regions of activity in the MTL identified in the EO, LO and EandL ROI maps. As noted in the analysis section, the process of creating the EO, LO, and EandL ROI maps could result in clusters with volumes of less than 300 vmul from ROI maps that had been previously thresholded at $p < .05$ corrected for multiple comparisons and therefore, since the anatomical regions within the MTL are also relatively small, I included any regions exceeding 100 vmul (approximately one voxel in original acquisition data) in the reported results.

In addition to the relatively small anatomical regions within the MTL, analysis and interpretation of MTL activations is complicated by several additional factors. First, encoding and retrieval are confounded in that every trial is both an encoding and a retrieval trial (Stark & Okado, 2003). Therefore, if encoding and retrieval are supported by the same neural tissue, then activation specific to these two tasks may not be identified. Second, the medial temporal lobe is particularly susceptible to elevated baselines relative to task conditions when the baseline condition does not require significant effort (Flanery, Law, & Stark, 2004; Law et al., Accepted pending revision; Newman et al., 2001; Stark & Squire, 2001). These elevated baseline levels can then result in either no differential activation or a relative deactivation of medial temporal areas for task relative to baseline conditions. In this second case, as in the case of relative activations, it is the pattern of activation relative to the other experimental conditions – increasing or decreasing - that is interpreted. Finally, due to its location, the MTL has a lower signal-to-noise ratio than other regions of cortex and therefore, statistically significant results are often elusive. However, even given these limitations in FMRI, the results of the ANOVA for areas of the MTL based on the EO, LO and EandL ROIs also appear to support the strong version of the rules-to-exemplars hypothesis. The results of this analysis are reported in Table 10.

The medial temporal activations reported in Table 10 show a relative deactivation of the medial temporal lobe relative to baseline during performance of the categorization LEARN and TEST blocks. These activations can be grouped into two patterns – two areas showing a relative decrease in activation with task experience – the type of pattern expected to be associated with

Table 10.

Medial Temporal Lobe Activation over the Course of Category Learning

vmul	LR	AP	IS	Hem	Brain Region	ROI map	F	p
166	24	-22	-12	R	Parahippocampal Gyrus	LO	1.033	.408
103	-15	-3	-13	L	Parahippocampal Gyrus	LO	2.402	.021
128	-18	-17	-15	L	Parahippocampal Gyrus	LO	1.752	.097
171	-22	-22	-14	L	Parahippocampal Gyrus	EandL	2.254	.030
498	-23	-20	-14	L	Parahippocampal Gyrus	EO	1.842	.079
266	-26	-36	-4	L	Hippocampus	EO	1.340	.231
167	-20	-41	-5	L	Parahippocampal Gyrus	EO	2.461	.018

R = Right (+), L = Left (-); A = Anterior (+), P = Posterior (-); S = Superior (+), I = Inferior (-); Hem = Hemisphere; vmul = volume; Effect = Factor with Significant Effect; F = value of F-statistic for the main effect of RUN; p = p-value associated with F statistic; LO = Late Only; EO = Early Only; EandL = Early and Late.

exemplar encoding and decreases in exemplar novelty (Kirwan & Stark, in press; Martin, 1999) – and three areas showing a relative increase in activation with task experience – the type of pattern expected to be associated with exemplar retrieval as memory strength increases (Flanery et al., 2004; Law et al., Accepted pending revision). The left hemisphere activations listed in Table 10 are located along an anterior-posterior hippocampal axis with the two anterior activations in the left panel (one significant, one approaching significance) of the figure showing a pattern of decreasing activation and the three posterior activations in the right panel (one significant, one approaching significance, one not significant) showing a pattern of increasing activation.

Evidence of an anterior-posterior axis associated with encoding and retrieval processes has previously been suggested by a meta-analysis completed by Tulving and colleagues (Tulving, Habib, Nyberg, Lepage, & McIntosh, 1999) and demonstrated empirically by Gabrielli, et al. (Gabrieli et al., 1997). However, the data on this issue remain open to debate (Schacter & Wagner, 1999). In the case of the present data, the reported results from the seven HC ROIs after conjunction analysis are over relatively small ROIs (498 to 103 vmul) with only 3 of the 7 meeting criteria for statistical significance ($p < .05$) and 2 of the 7 approaching significance ($p < .10$). It may be the specificity of these smaller regions that allows this pattern to emerge since larger ROIs would group together two different activation patterns resulting in a pattern that looks nonspecific to either encoding or retrieval. In fact, the HC region located between the left hemisphere anterior and posterior HC activations, although having a statistically significant main effect of RUN, showed a pattern of activation that was not specific

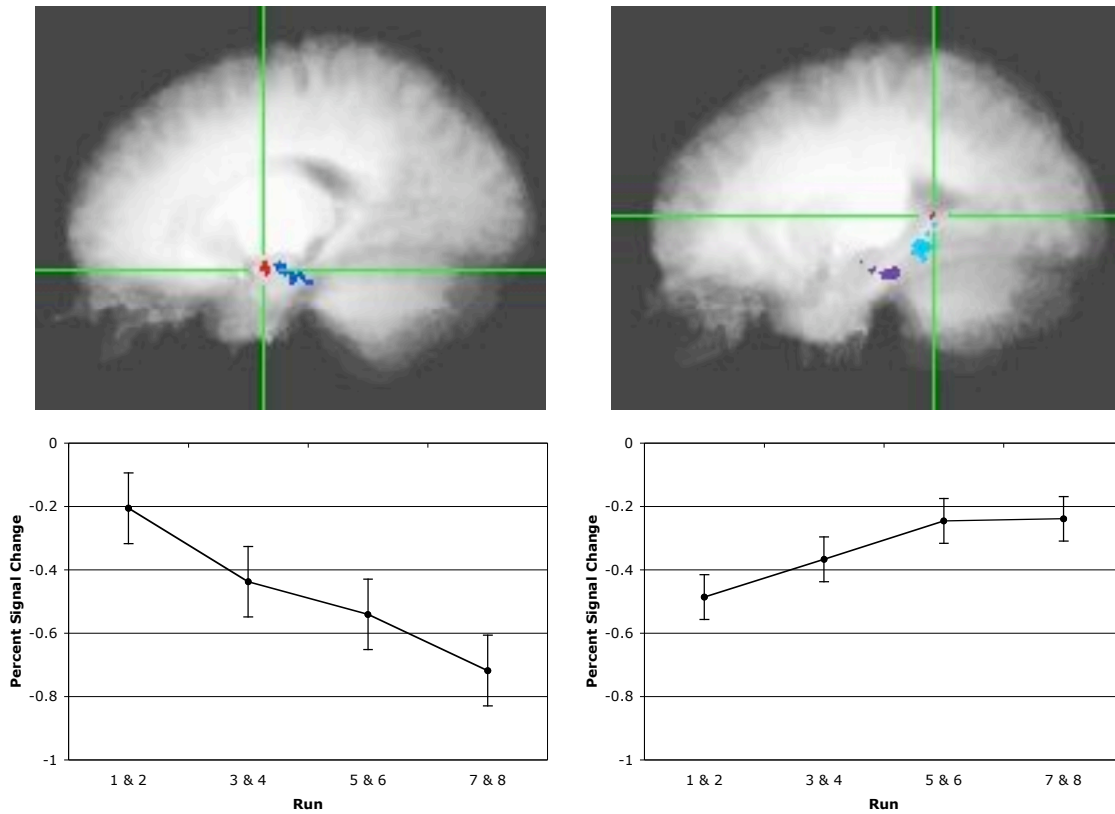


Figure 9. Activation of left hippocampal complex. The panel on the left shows the two anterior activations from the LO ROI map and the right panel shows the three posterior activations from the EO ROI map. The graphs depict the average activation for each quartile for the most anterior (left graph) and most posterior (right graph) regions. Activations for both of these regions showed a statistically significant effect of RUN.

to either encoding or retrieval. This could be the result of either an HC region that is equally active with both encoding and retrieval or the effect of averaging an HC region that includes separate areas – one active with encoding and one active with retrieval. Additional reasons why this pattern may have emerged in the data are that I collected data over the entire learning sequence, the data were analyzed without collapsing across these different learning epochs, and the method used to identifying ROIs based on differential activation during either early or late ROIs might be better able to identify these contrasting patterns.

An additional pattern seen in the results is that 6 of the 7 MTL regions identified by the EO, LO, and EandL ROI maps were located in the left hemisphere. While no conclusions can be drawn from this data with respect to hemispheric specialization, abstract processing has

previously been associated with left hemisphere processing and specific object processing has been associated with right hemisphere processing (Marsolek, 1995, 1999). However, it has been suggested that the left hemisphere advantage reported by Marsolek for abstract objects may be attributable to a left hemisphere advantage for semantic processing (Curby, Hayward, & Gauthier, 2004).

Anatomical Atlas Basal Ganglia ROIs

Since the fronto-striatal system had been found to be differentially active with task differences in Experiments 1 and 2, I also completed an analysis of the basal ganglia regions using ROIs based on the Talairach Atlas. However, none of the regions defined in this manner were found to have significant main effects of RUN or TASK.

Attention versus Rules

Two possible alternatives to the rules-to-exemplars hypothesis are that the observed behavioral shift may be accounted for by a shift in selective attention between features (Johansen & Palmeri, 2002; Nosofsky & Johansen, 2000) or shifts in selective attention between rules (Ericksen & Kruschke, 1998). The current data do not speak to this issue. The evidence of process dissociation I have reported in the present study, although consistent with a rules-to-exemplars shift, does not exclude the possibility that mechanisms of selective attention are supporting categorization performance. Rules and attentional processes may be the same or interdependent processes since disrupting attentional processes appears to interfere more with categorization tasks that depend on rule application (Ashby, Maddox, & Bohil, 2002; Maddox, Ashby, Ing, & Pickering, 2004; Waldron & Ashby, 2001).

Due to its ubiquitous nature, attention must almost certainly play a part in the categorization process. However, if shifts in selective attention are part of this process, any theory proposing shifts in attention supporting this behavioral shift will also need to accommodate evidence from the current study which clearly supports a pattern of early neural activity supporting rule application and exemplar encoding combined with late neural activity supporting exemplar retrieval. Taken together, the results of these analyses support the conclusion that there is a shift in categorization systems or representations from a rule-based system supporting categorization judgments during early epochs of the task to an exemplar (or prototype) based system supporting categorization judgements during late epochs of the task.

CHAPTER V

GENERAL DISCUSSION

The experiments in this dissertation have explored the question – What are the neural correlates of explicit categorization? Analysis and Discussion of Experiments 1, 2, and 3 focused on identifying neural correlates differentially active with task manipulations that had previously been associated with multiple- or hybrid systems of categorization. These experiments utilized fMRI to identify areas of differential neural activity which, if found, would provide a converging source of evidence supporting a multiple-categorization systems view of categorization. And, in general, it would appear that there is evidence of differential activation across a variety of category learning and testing conditions. However, it should be noted that finding differential activation in an fMRI study does not necessarily mean that there are multiple systems supporting a cognitive process. Whereas multiple systems are inferred in behavioral studies from double dissociations, it is not possible to definitively identify a separate system using whole brain fMRI in normal subjects because differential activation may be evidence of *functional segregation*, where processing depends on one and only one neural area, *functional integration*, where processing depends on two or more areas, or *functional degeneracy*, in which several sufficient systems can support processing (Price & Friston, 2002). To find affirmative evidence of a separate system utilizing fMRI, neuroimaging data from both normal and brain damaged subjects with specific lesions to each area showing functional differentiation must be collected - a task which is beyond the scope of this dissertation.

Thus, the results of Experiments 1, 2, and 3, may suggest, but do not prove, the existence of multiple categorization systems. However, another aspect of the data collected and analyzed in Experiments 1, 2, and 3, suggests that categorization is subserved by a stable system of neural regions that are active with categorization regardless of task differences. In the current chapter I examine the neural regions in Experiments 1, 2, and 3 that were found to be differentially active with the categorization relative to the control conditions. The coordinates for the voxel of maximum intensity for these regions were reported in Tables 3, 6, and 8 for Experiments 1, 2, and 3, respectively, and used for the differential analyses presented for those experiments.

Table 11
Areas of Activation Common across Experiments 1, 2, and 3

vmul	LR	AP	SI	Hem	Gyrus	BA
<u>Frontal</u>						
3860	-3	53	4	Left	Medial Frontal Gyrus	10
774	41	36	17	Right	Middle Frontal Gyrus	46/10
4206	2	20	39	Right	Cingulate Gyrus	32
600	-31	20	1	Left	Insula/Inferior Frontal Gyrus	13/47/45
1840	34	19	-1	Right	Inferior Frontal Gyrus	47/13
6327	38	8	38	Right	Middle Frontal Gyrus	9
843	-42	7	27	Left	Inferior Frontal Gyrus	9
394	-27	1	49	Left	Middle Frontal Gyrus/SFG	6
<u>Limbic</u>						
1717	13	-5	10	Right	Thalamus/Caudate	-
602	-13	-8	11	Left	Thalamus/Caudate	-
<u>Temporal/Parietal/Occipital</u>						
7253	-3	-36	37	Left	Cingulate Gyrus	31
19676 ^a	-29	-52	46	Left	Inferior Parietal Lobule	Var
2112	-47	-59	31	Left	Angular Gyrus	39
30888 ^a	27	-60	37	Right	Precuneus	Var
8815	-37	-71	2	Left	Middle Occipital Gyrus	Var

To identify areas associated with categorization relative to control commonly active across the three experiments, the thresholded maps for the categorization relative to the control conditions for each experiment ($p < .05$, corrected for multiple comparisons) were blurred with a 3 mm filter, averaged and then thresholded to include only areas of maximal overlap across the three experiments. vmul = volume in mm³; R = Right (+), L = Left (-); CM = Center of Mass, A = Anterior(+), P = Posterior(-), S = Superior (+), I = Inferior (-), Hem = Hemisphere; BA = Brodmann Area.

^a – large regions extending over multiple neural regions were listed based on the location of the voxel at the center of mass

To identify areas associated with categorization relative to control commonly active across the three experiments, I took the thresholded maps for the categorization relative to the control conditions for each experiment ($p < .05$, corrected for multiple comparisons), applied a 3 mm blurring filter to each map, averaged the maps, and then thresholded the maps to include only areas of maximal overlap across the three experiments. Using this procedure, I identified fifteen areas of neural activation common to all three experiments. The Talairach labels and coordinates for the centers of mass for these areas are listed in Table 11 and illustrated in Figure 10.

This analysis reveals multiple neural regions that are active when subjects perform categorization but not when they respond to visually similar stimuli. These fifteen regions

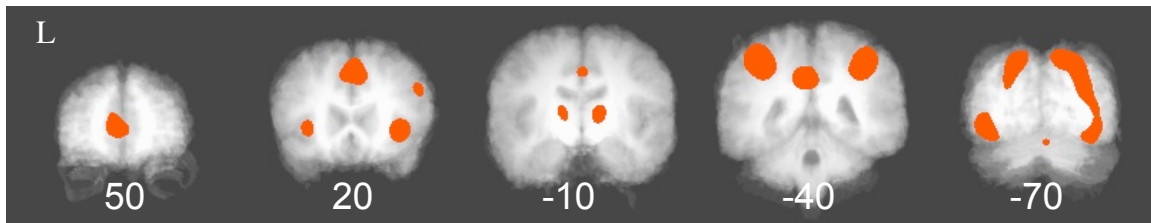


Figure 10. Neural areas active with categorization relative to control tasks common across Experiments 1, 2 and 3. Five coronal slices spaced 30 mm apart starting at 50 mm anterior to the anterior commissure (AC) and ending 70 mm posterior to the AC. Slices are presented with the left (L) hemisphere on the left side of the image. The Talairach coordinates for the center of mass for each region are reported in Table 11.

showing sustained activity during the categorization tasks relative to the control tasks demonstrate that categorization is also supported by the sustained activity by this group of neural regions. The three occipital areas identified are generally associated with perceptual processing of visual stimuli (Buckner et al., 1998; Van Essen, Drury, Joshi, & Miller, 1998; Van Essen et al., 2001), the frontal regions with working memory, cognitive control, and rule application (Braver et al., 1997; Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998; Elliott et al., 1999; Fletcher & Henson, 2001; Goldman-Rakic, 1995; Miller, 2000; Miller & Cohen, 2001; Wallis et al., 2001; Wallis & Miller, 2003), the cingulate regions with cognitive control and error detection/monitoring (Carter et al., 1998; Elliott & Dolan, 1998; Peterson et al., 1999), parietal regions with attention (Colby & Goldberg, 1999; Corbetta et al., 1998; Culham & Kanwisher, 2001; Desimone & Duncan, 1995; Kanwisher & Wojciulik, 2000; Shafritz et al., 2002; Shipp, 2004) and the caudate/basal ganglia with stimulus/response mapping (Goldman-Rakic, 1995; Graybiel, 1995; Graybiel, Aosaki, Flaherty, & Kimura, 1994; Gurney et al., 2001; Hayes, Davidson, Keele, & Rafal, 1998; Joel et al., 2002; Packard & Knowlton, 2002). And, these same areas have also been found differentially active in fMRI studies of categorization in addition to the three studies reported in this dissertation (Aizenstein et al., 2000; Poldrack et al., 2001; Poldrack et al., 1999; Reber et al., 2003; Reber et al., 1998a, 1998b; Reber et al., 2002; Seger & Cincotta, 2002; Seger et al., 2000; Vogels et al., 2002).

If discussion had been limited to the results of the differential analyses reported in Chapters III and IV, I may have interpreted the evidence reported in those chapters as supporting multiple-categorization systems since the areas of reported differential activation did not overlap across these studies. Conversely, had I limited the analysis to the results reported in the current

Chapter, I may have interpreted the evidence in terms of a single-categorization system since the areas of reported differential activation showed significant amounts of overlap.

Based on a review of both implicit and explicit categorization studies utilizing a variety of paradigms and experimental methods, Keri (2003) has proposed an integrative model of category learning in which three category learning subsystems – subserved by the medial temporal lobe, the basal ganglia, and sensory neocortex – are mediated by abstract rules, attentional allocation, and decisional processes – subserved by the lateral prefrontal cortex and anterior cingulate cortex. As I will discuss in the following paragraphs, this model would appear to fit well with the differential results reported for Experiments 1, 2, and 3 as well as with the overall results reported in Table 11.

Learning Subsystems

Basal Ganglia

In accord with Keri's model in which the basal ganglia are an important category learning subsystem, the combined results clearly show bilateral activation of portions of the caudate/thalamus for categorization relative to control tasks. The coronal image in the center panel of Figure 10 shows a slice of cortex that illustrates an area of bilateral activation of the caudate/thalamus. Keri's (2003) model proposes that the basal ganglia is associated with stimulus-response and stimulus-reward mapping - an analysis that is supported by the differential results reported for Experiments 1 and 2. In Experiment 1, although the basal ganglia were differentially active for categorization relative to control conditions, I did not find a statistically significant difference in activation within the basal ganglia between early deterministic categorization and early probabilistic categorization.

In Experiment 2, I again found the basal ganglia active for categorization relative to control conditions. In addition, differential activation was reported in the right thalamus with respect to the learning task (feedback versus exposure) and an interaction of learning (feedback versus exposure) and task phase (learning versus testing) was found in the left thalamus. But again I did not find differences within the basal ganglia during learning with respect to the type of category feedback structure (probabilistic versus deterministic) that was being used. Thus, across these two experiments the basal ganglia were active when subjects were associating a stimulus – regardless of the category feedback structure.

Poldrack and colleagues (1999, 2001) have previously reported basal ganglia activity with the probabilistic card task and medial temporal activity with a paired associates version of the same task (similar to Experiment 2). However, in these studies, the probabilistic task was learned in a feedback paradigm while the paired-associate task was learned in an exposure paradigm. Thus, the difference in activation could be associated with the type of category structure being used in the two conditions or the feedback versus exposure learning use in the two conditions. The results of Experiment 2 control for this confound and do not report differential activation between category structure conditions when controlled for the type of learning condition, again supporting Keri's (2003) proposal that the basal ganglia are associated with stimulus response and/or reward mapping rather than with a particular type of category structure.

Medial Temporal Lobe

Of the three category learning subsystems identified by Keri, the medial temporal system is the only system for which I do not find evidence in the combined results reported in Table 11. However, anterior, ventral anterior, and medial portions of the medial temporal lobe are highly susceptible to imaging artifacts such as signal drop-out due to their close proximity to the sinus cavities. In addition, the signal-to-noise ratio in these areas is very low and Experiment 1 was scanned using a 1.5 Tesla MRI scanner. In addition, the hippocampus and related cortical regions are relatively small and variable between subjects. Most studies that find activation within this region use ROI analyses based on a small volume correction and/or use specialized alignment techniques based on anatomical markers. Although I did not find evidence of medial temporal activity in the combined results reported in Table 11, I did identify neural activity in the hippocampal complex associated with encoding and/or retrieval in Experiment 3.

Sensory Neocortex

Keri (2003) identifies a sensory neocortex learning subsystem that is sensitive to stimulus similarity and frequency. Keri posits this system based on results from categorization studies in which learning is incidental/implicit such as the series of dot-pattern studies completed by Reber and colleagues (Reber & Squire, 1999; Reber et al., 1998a, 1998b). Since the experiments in this dissertation utilized explicit training and testing, differential activation within sensory neocortex

was not necessarily an expected outcome in the differential analyses. However, implicit/incidental learning may still occur in the presence of explicit learning and therefore, differential activation of sensory neocortex might still occur. In fact, I found widespread activation of occipital neocortex for categorization relative to control in the combined results. The coronal image on the far right of Figure 10 shows a slice of cortex that includes bilateral activation of both parietal and occipital cortex. Since all of the control conditions included stimuli that were matched with the categorical condition stimuli on both number and length of exposure, the differential activation cannot be attributed to perceptual priming. Thus, across the three experiments in this dissertation, sensory neocortex showed greater activation with respect to blocks of categorical stimuli than noncategorical stimuli. However, without further experiments, it is unclear what the significance of this increased activation in sensory neocortex may be. Lamme has proposed that conscious processing of stimuli depends on feedback from higher level visual and association areas to lower level sensory neocortex (Lamme & Roelfsema, 2000). It may be this feedback processing in occipital and parietal cortex that is occurring during blocks of categorization - where the stimuli are relevant to the experimental task – relative to the control stimuli – where the stimuli have no experimental relevance to the participant.

Mediation by Orbitofrontal, Lateral Prefrontal, and Anterior Cingulate Cortex

In addition to the three learning subsystems, Keri (2003) identifies networks of neural regions including the orbitofrontal, lateral prefrontal, and anterior cingulate cortex that support abstract rule generation, top-down attentional control, and decision making processes which he characterizes as mediating the three learning subsystems. Again, evidence of these cortical networks is also found in the differential and combined results reported above. For instance, in Experiment 3 I identified neural regions associated with rule application in previous functional imaging studies and then examined functional activation within these areas while subjects learned ill-defined categories. In a subset of these regions activation during early phases of learning was significantly different than activation during late phases as would be expected with a shift away from rule use during later epochs of the task. In the group results, there were also significant portions of the cingulate cortex commonly active across the three experiments with categorization relative to control. Neuroimaging studies have associated the cingulate with functions such as error monitoring, cognitive control, response selection, and attention (Carter et

al., 1998; Elliott & Dolan, 1998; Peterson et al., 1999) making this region a likely candidate for the mediation among the learning subsystems suggested by Keri. Finally portions of a fronto-parietal attentional network (Shipp, 2004) are shown as active in the combined results as would also be predicted by Keri's model of category learning.

Conclusion

It should be clear from the analyses presented in Chapters III, IV, and V that categorization is a dynamic, whole brain process that is responsive to changes in the learning environment. Had I limited the analysis to statistical comparisons of activation during categorization relative to activation during perceptual control tasks, I may have concluded that categorization was well specified by a single-system model since a common network of areas was active during early and late learning as well as across different types of learning paradigms. Similarly, had I limited the analysis to statistical comparisons of the differential activation between categorization tasks, I may have concluded that categorization was well specified by a multiple system model since these analyses result in areas of differential activation that do not overlap. The answer most likely lies somewhere between these extremes – similar to a component processes view of categorization (Palmeri & Flanery, 2002; Roediger, Buckner, & McDermott, 1999; Witherspoon & Moscovitch, 1989) which, as described by Witherspoon and Moscovitch is :

'based on the assumption that performance on each task requires the operation of many components, some of which are common to tasks and some of which are not. Performances on each task may be independent from each other to the extent that their components differ (or the information they use is different), leaving open the possibility that some components (or types of information) may be more critical in this regard than others.' (Witherspoon & Moscovitch, 1989, p. 89)

At the outset of this dissertation, I stated that categorization is a fundamental cognitive process that allows us to fluidly interact with our environment. If we take a step back from the experimental data that have been collected over the years, one can note that no one has identified a subject population that is completely at chance at all forms of categorization. Even with PARK

subjects, categorization performance is impaired – but still above chance – on only a particular form or strategy of category learning. From this one piece of observational datum, one could have concluded that categorization must involve a set of neural regions that are combined in such a way as to be fairly robust in the face of injury – something similar to a group of component processes.

If categorization is best described as relying on a group of component processes, it has implications for both past and future research. For instance, if categorization is supported by component processes, it becomes more challenging to instantiate an overall theory of categorization and then test models of this theory quantitatively. Many questions will need to be answered with respect to each of the component categorization processes. For instance, with respect to the use of rules during early learning, is the use of rules a required component of category learning or just one of several possible methods or strategies that can be used to acquire the category? If there are several ways in which the category can be acquired, do they work in parallel or is there top-down control of which method is used? Are there aspects of the learning environment that can evoke one method over another? Do neural differences during learning result in neural differences during transfer? The experiments in this dissertation provide only a first step in addressing these questions.

In the past, cognitive psychologists had the output of a single black box – the brain – with which to understand cognitive processes. Manipulation of experimental variables resulted in the change of one or two behavioral markers that could then be interpreted within the framework of various theories. With the advent of functional neuroimaging, the single black box has grown to over 70,000 black boxes – the number of voxels in an average whole brain neuroimaging study – that change in varying ways with the same manipulation of an experimental variable providing us with a rich set of data with which to understand cognitive processes. But, there is no such thing as a free lunch (Friedman, 1974; Heinlein, 1969). The cost of these 70,000 observations is greatly increased complexity of the results to be interpreted. But the benefit of this increased complexity may be the ability to reach our ultimate goal – understanding the relationship between the mind and the brain.

APPENDIX A

Stimuli - Experiment 2

Stimuli for experiment 2 were created by combining eight nameable objects and eight possible configurations. Using these objects and configurations, a total of 64 different cards can be generated for use in the various conditions of experiment 2. The same set of 64 cards were used for the control condition with the exception that the objects were red objects on white cards with red borders.

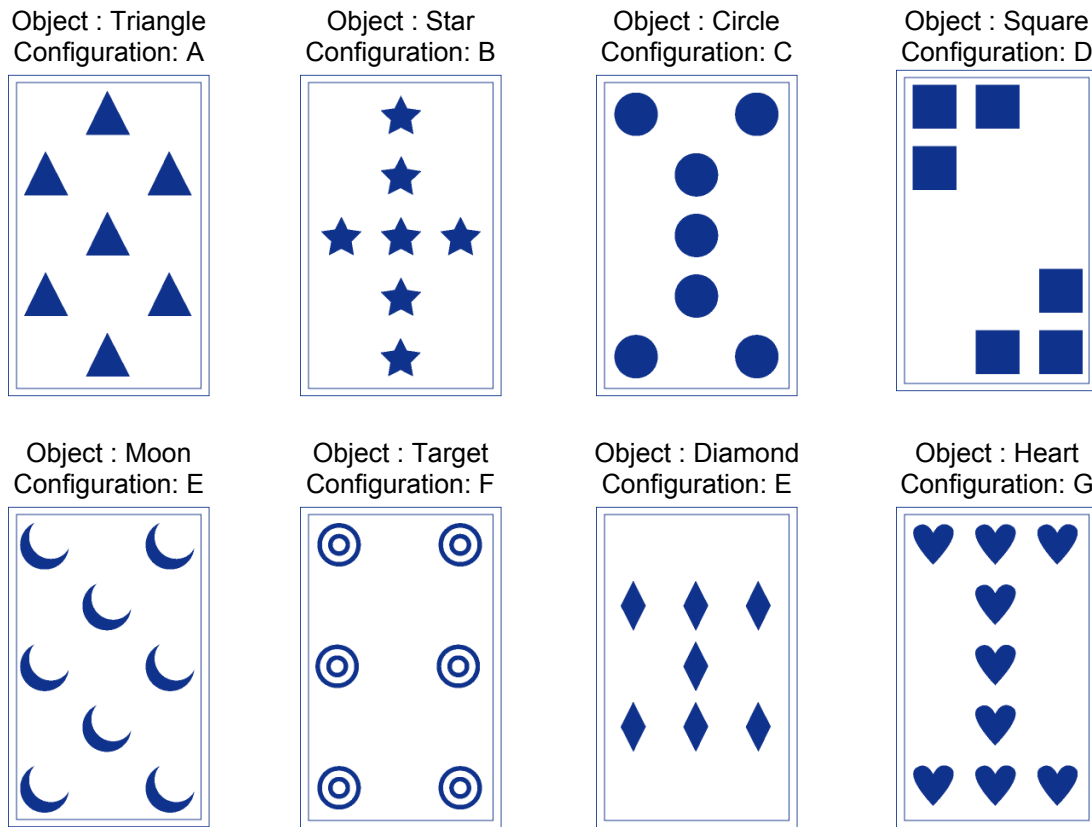


Figure 11. Card stimuli used in Experiment 2.

APPENDIX B

Materials – Experiment 3

Stimulus Construction and Norming

Stimuli for Experiment 3 consisted of four sets of stimuli - Rockets, Lamps, Sailboats and, Neckties. The rocket stimuli were modeled on stimuli previously used by Nosofsky et al. (Nosofsky & Palmeri, 1998; Nosofsky, Palmeri et al., 1994) and the lamp stimuli were modeled on stimuli previously used by Lamberts (Lamberts & Freeman, 1999). The sailboat and necktie stimuli were newly created for use in the experiment. All of the stimuli had four binary-valued features. To assure that the salience of the stimulus dimensions was roughly equal we collected similarity ratings for pairs of stimuli that varied by 0, 1 or 2 features (similar to (Minda & Smith, 2002)) as described in the following paragraph. The goal of this testing was to create features that, when changed, resulted in the same similarity ratings.

For each trial, subjects were presented two stimuli, sequentially, for 2000 ms each. After the second stimulus was erased from the screen, subjects were asked to rate how similar the stimuli were to each other on a scale of 1 (Very Alike) to 5 (Very Different). Subjects were allowed 3000 ms to use the numeric keypad to record their judgements. In written and verbal instructions subjects were told "*when making your similarity judgment, it is important you attend to ALL of the features of the objects rather than any single feature.*" All ratings were made within category – that is lamps were only compared to lamps.

Data were then analyzed for the comparison conditions in which the stimuli varied by only one dimension. If the average similarity ratings for these six comparisons were found to be statistically significant (as confirmed by ANOVA), the features would be adjusted to increase or decrease the salience of the particular feature or features that based on the subject ratings.

Table 12. Rocket Dimensions

	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>
	Wing	Tail	Nose	Decoration
binary value 0	triangular	jagged	square	stripes
binary value 1	rectangular	boxed	pointed	triangles

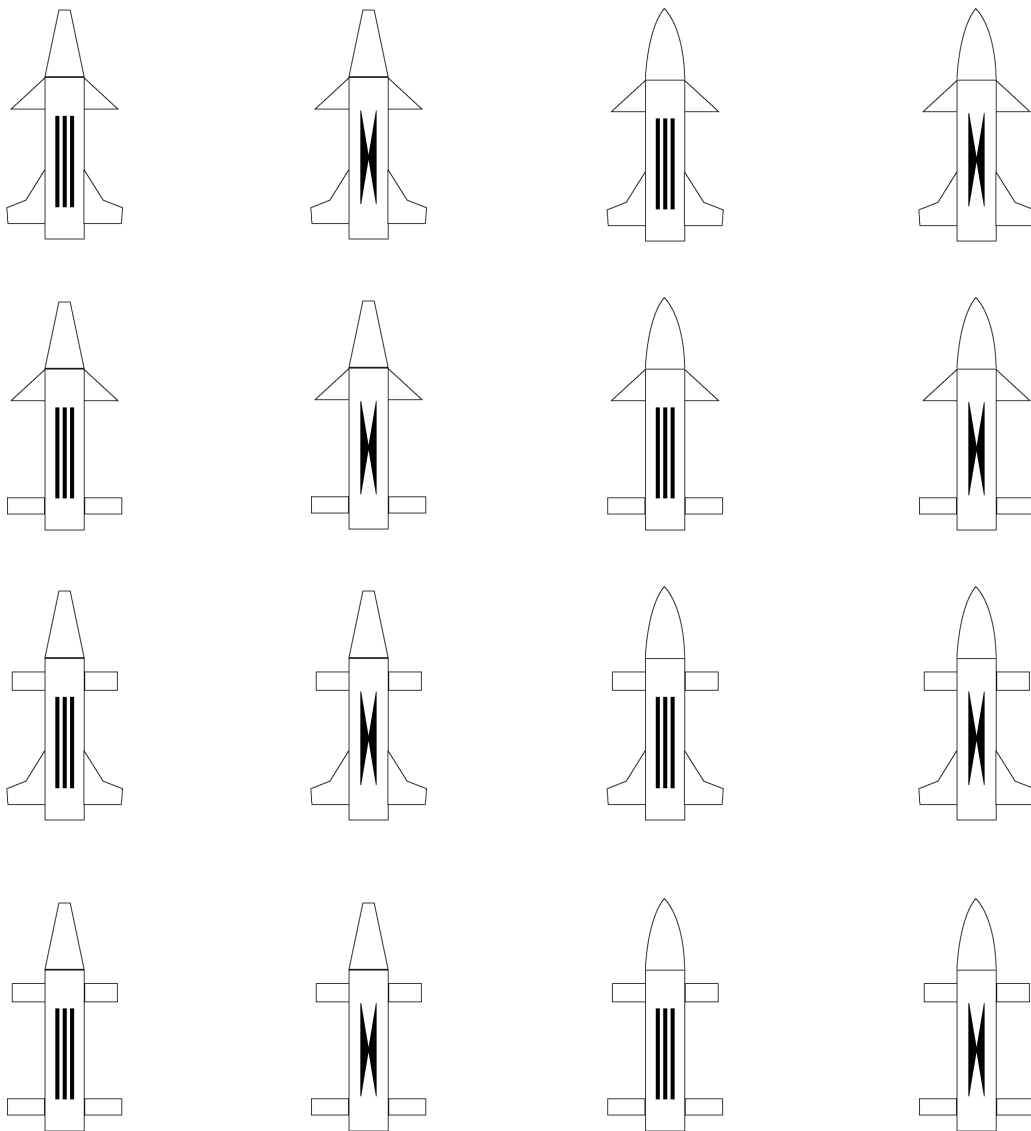


Figure 12. Rocket stimuli used in Experiment 3.

Table 13. Lamp Dimensions

	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>
	Base	Upright	Shade	Finial
binary value 0	smooth	thick	wide	spherical
binary value 1	stacked	thin	thin	rectangular

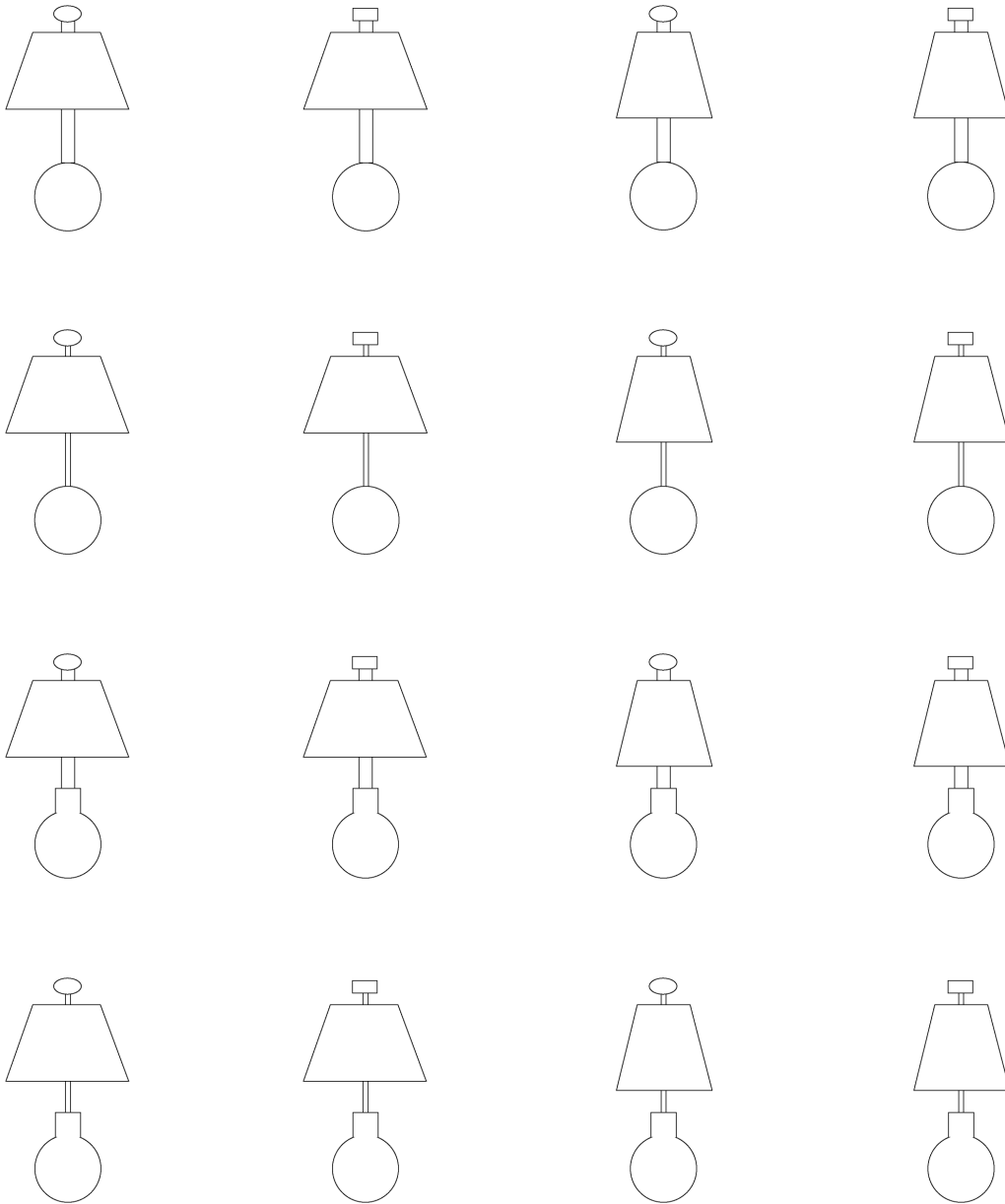


Figure 13. Lamp stimuli used in Experiment 3.

Table 14. Sailboat Dimensions

	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>
	Hull	Flag	Porthole	Sail
binary value 0	angular	rectangular	round	long
binary value 1	rounded	triangular	diamond	short

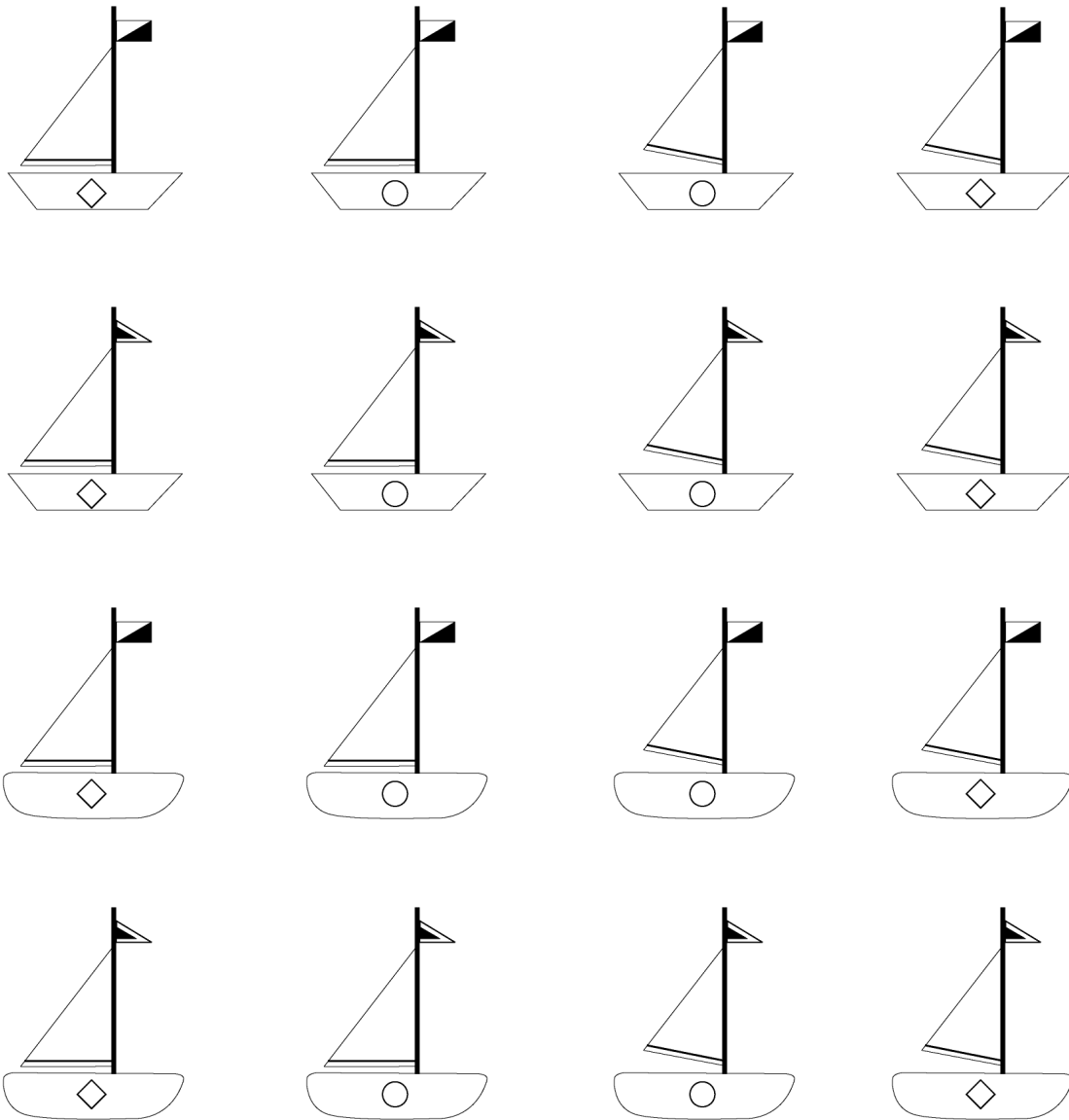


Figure 14. Sailboat stimuli used in Experiment 3.

Table 15. Necktie Dimensions

	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>
	Bow Stripes	Pattern	Width	Bow Shape
binary value 0	vertical	hexagon	thin	hexagon
binary value 1	horizontal	circle	fat	square

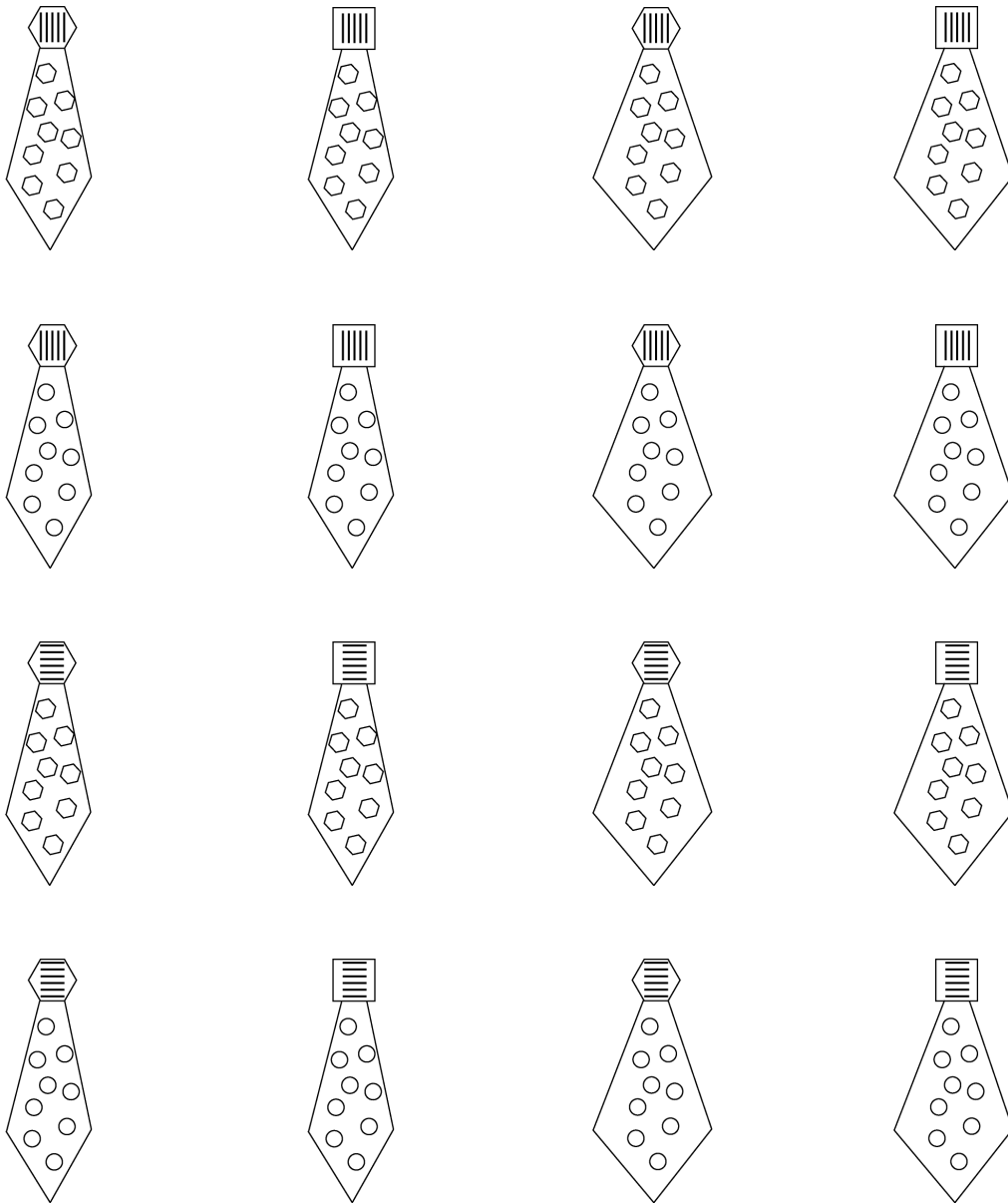


Figure 15. Necktie stimuli used in Experiment 3.

REFERENCES

- Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences*, 22, 425-489.
- 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.
- Alvarez, P., Zola-Morgan, S., & Squire, L. R. (1995). Damage limited to the hippocampal region produces long-lasting memory impairment in monkeys. *The Journal of Neuroscience*, 15(5), 3796-3807.
- Amaral, D. G. (1999). Introduction: What is where in the medial temporal lobe? *Hippocampus*, 9, 1-6.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98(3), 409-429.
- 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(3), 442-481.
- Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(1), 33-53.
- Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, 37, 372-400.
- Ashby, F. G., Maddox, W. T., & 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., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review*, 6(3), 363-378.
- Baddeley, A., Vargha-Khadem, F., & Mishkin, M. (2001). Preserved recognition in a case of developmental amnesia: Implications for the acquisition of semantic memory? *Journal of Cognitive Neuroscience*, 13(3), 357-369.

- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*, *5*, 49-62.
- Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*, *39*, 713-726.
- Broman, M., Rose, A. L., Hotson, G., & Casey, C. M. (1997). Severe anterograde amnesia with onset in childhood as a result of anoxic encephalopathy. *Brain*, *120*, 417-433.
- Bruner, J. S., Goodnow, J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., et al. (1998). Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron*, *20*, 285-296.
- Bunge, S., Kahn, I., Wallis, J. D., Miller, E. K., & Wagner, A. D. (2003). Neural circuits subserving the retrieval and maintenance of abstract rules. *Journal of Neurophysiology*, *90*, 3419-3428.
- Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, *12*(1), 1-47.
- Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D. C., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747-749.
- Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, *22*, 319-349.
- Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., et al. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*, 761-773.
- Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science*, *279*, 1347-1351.
- Cox, R. W. (1996). AFNI: Software for the analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162-173.
- Culham, J. C., & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology*, *11*, 157-163.

- Curby, K. M., Hayward, W. G., & Gauthier, I. (2004). Laterality effects in the recognition of depth-rotated novel objects. *Cognitive, Affective, & Behavioral Neuroscience*, 4(1), 100-111.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193-222.
- Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Science, USA*, 98(24), 13763-13768.
- Elliott, R., & Dolan, R. J. (1998). Activation of different anterior cingulate foci in association with hypothesis testing and response selection. *NeuroImage*, 8, 17-29.
- Elliott, R., & Dolan, R. J. (1999). Differential neural responses during performance of matching and nonmatching to sample tasks at two delay intervals. *The Journal of Neuroscience*, 19(12), 5066-5073.
- Elliott, R., Rees, G., & Dolan, R. J. (1999). Ventromedial prefrontal cortex mediates guessing. *Neuropsychologia*, 37, 403-411.
- Ericksen, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127(2), 107-140.
- Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, 18, 500-549.
- Filoteo, J. V., Maddox, W. T., & Davis, J. D. (2001). Quantitative modeling of category learning in amnesic patients. *Journal of the International Neuropsychological Society*, 7, 1-19.
- Flanery, M. A., Law, J. R., & Stark, C. E. L. (2004). *When negative isn't negative*. Paper presented at the Society for Neuroscience, San Diego, CA.
- Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain*, 124, 849-881.
- Friedman, M. (1974). *There's No Such Thing As a Free Lunch*. LaSalle, Ill: Open Court.
- Gabrieli, J. D. E., Brewer, J. B., Desmond, J. E., & Glover, G. H. (1997). Separate neural bases of two fundamental memory processes in the human medial temporal lobe. *Science*, 276, 264-266.
- Gluck, M. A., Oliver, L. M., & Myers, C. E. (1996). Late-training amnesic deficits in probabilistic category learning: A neurocomputational analysis. *Learning & Memory*, 3, 326-340.

- 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). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F. Plum (Ed.), *Handbook of Physiology: The Nervous System*.
- Goldman-Rakic, P. S. (1995). Toward a circuit model of working memory and the guidance of voluntary motor action. In J. C. Houk, J. L. Davis & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia*. Cambridge, Massachusetts; London, England: The MIT Press.
- Graybiel, A. M. (1995). Building action repertoires: Memory and learning functions of the basal ganglia. *Current Opinion in Neurobiology*, 5, 733-741.
- Graybiel, A. M., Aosaki, T., Flaherty, A. W., & Kimura, M. (1994). The basal ganglia and adaptive motor control. *Science*, 265, 1826-1831.
- Gurney, K., Prescott, T. J., & Redgrave, P. (2001). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological Cybernetics*, 84, 401-410.
- Hayes, A. E., Davidson, M. C., Keele, S. W., & Rafal, R. D. (1998). Toward a functional analysis of the basal ganglia. *Journal of Cognitive Neuroscience*, 10(2), 178-198.
- Heinlein, R. (1969). *The Moon is a Harsh Mistress*. London: New English Library.
- Hintzman, D. L. (1986). "Schema Abstraction" in a multiple-trace memory model. *Psychological Review*, 93(4), 411-428.
- Homa, D. (1978). Abstraction of ill-defined form. *Journal of Experimental Psychology: Human Learning and Memory*, 4(5), 407-416.
- Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory*, 2(3), 322-330.
- Hopkins, R. O., Myers, C. E., Shohamy, D., Grossman, S., & Gluck, M. A. (2004). Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia*, 42, 524-535.
- Joel, D., Niv, Y., & Ruppin, E. (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Networks*, 15, 535-547.
- Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, 45, 482-553.

- Johnstone, T., & Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology*, 42, 61-112.
- Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews Neuroscience*, 1, 91-100.
- Kirwan, B. C., & Stark, C. E. L. (in press). Medial temporal lobe activation during encoding and retrieval of novel face-name pairs. *Hippocampus*.
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, 273, 1399-1402.
- Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, 262, 1747-1749.
- Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, 1, 106-120.
- Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, 51, 481-537.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99(1), 22-44.
- Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, 107(2), 227-260.
- Lamberts, K., & Freeman, R. P. J. (1999). Categorization of briefly presented objects. *Psychological Research*, 62(2-3), 107-117.
- Lamme, V. A. F. (2000). Neural mechanisms of visual awareness: A linking proposition. *Brain and Mind*, 1, 385-406.
- Lamme, V. A. F. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience*, 14(7), 1044-1053.
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience*, 23(11), 571-579.
- Law, J. R., Flanery, M. A., Wirth, S., Yanike, M., Suzuki, W. A., Smith, A. C., et al. (Accepted pending revision). fMRI activity during the gradual acquisition and expression of paired-associate memory. *Journal of Neuroscience*.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95(4), 492-527.

- Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning-curves - A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 883-914.
- Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, 109(2), 376-400.
- 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.
- Marsolek, C. J. (1995). Abstract visual-form representation in the left cerebral hemisphere. *Journal of Experimental Psychology: Human Perception and Performance*, 21(2), 375-386.
- Marsolek, C. J. (1999). Dissociable neural subsystems underlie abstract and specific object recognition. *Psychological Science*, 10(2), 111-118.
- Martin, A. (1999). Automatic activation of the medial temporal lobe during encoding: Lateralized influences of meaning and novelty. *Hippocampus*, 9, 62-70.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 103(3), 419-457.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85(3), 207-238.
- Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Natures Reviews: Neuroscience*, 1, 59-65.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167-202.
- Miller, E. K., Freedman, D. J., & Wallis, J. D. (2002). The prefrontal cortex: Categories, concepts and cognition. *Phil. Trans. R. Soc. Lond. B*, 357, 1123-1136.
- Milner, B. (1968). Preface: Material-specific and generalized memory loss. *Neuropsychologia*, 6, 175-179.
- Milner, B., Corkin, S., & Teuber, H.-L. (1968). Further analysis of the hippocampal amnesic syndrom: 14-year follow-up study of H.M. *Neuropsychologia*, 6, 216-234.

- 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.
- Newman, S. D., Twieg, D. B., & Carpenter, P. A. (2001). Baseline conditions and subtractive logic in neuroimaging. *Human Brain Mapping*, 14(4), 228-235.
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 104-114.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115(1), 39-57.
- Nosofsky, R. M. (1988a). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(4), 700-708.
- Nosofsky, R. M. (1988b). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14(1), 54-65.
- Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, 17(1), 3-27.
- Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, 22(3), 352-369.
- 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., & 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). Rule-plus-exception model of classification learning. *Psychological Review*, 101(1), 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(4), 247-255.
- O'Reilly, R. C., Noelle, D. D., Braver, T. S., & Cohen, J. D. (2002). Prefrontal cortex and dynamic categorization tasks: Representational organization and neuromodulatory control. *Cerebral Cortex*, 12, 246-257.

- O'Reilly, R. C., & Rudy, J. W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus*, *10*, 389-397.
- O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, *108*(2), 311-345.
- 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.
- Owen, A. M., Milner, B., Petrides, M., & Evans, A. C. (1996). A specific role for the right parahippocampal gyrus in the retrieval of object-location: A positron emission tomography study. *Journal of Cognitive Neuroscience*, *8*(6), 588-602.
- Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience*, *25*, 563-593.
- Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 324-354.
- Palmeri, T. J., & Flanery, M. A. (1999). Learning about categories in the absence of training: Profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychological Science*, *10*(6), 526-530.
- Palmeri, T. J., & Flanery, M. A. (2002). Memory systems and perceptual categorization. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 41): Academic Press.
- Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 548-568.
- Palmeri, T. J., & Nosofsky, R. M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *The Quarterly Journal of Experimental Psychology*, *54A*(1), 197-235.
- Parkin, A. J. (1997). *Memory and Amnesia: An Introduction* (2nd ed.). Oxford, UK: Blackwell.
- Patalano, A. L., Smith, E. E., Jonides, J., & Koeppel, R. A. (2001). PET evidence for multiple strategies of categorization. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(4), 360-370.
- Peterson, B. S., Skudlarski, P., Gatenby, J. C., Zhang, H., Anderson, A. W., & Gore, J. C. (1999). An fMRI study of stroop word-color interference: Evidence for cingulate subregions subserving multiple distributed attentional systems. *Biological Psychiatry*, *45*(10), 1237-1258.

- Pettersson, K. M., Elfgren, C., & Ingvar, M. (1999). Learning-related effects and functional neuroimaging. *Human Brain Mapping, 7*, 234-243.
- Petrides, M. (1996). Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Phil. Trans. R. Soc. Lond. B, 228*, 105-116.
- Petrides, M. (2002). The Mid-ventrolateral prefrontal cortex and active mnemonic retrieval. *Neurobiology of Learning and Memory, 78*, 528-538.
- Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6*(10), 13634-.
- Poldrack, R. A. (2000). Imaging brain plasticity: Conceptual and methodological issues - A theoretical review. *NeuroImage, 12*, 1-13.
- Poldrack, R. A., Clark, J., Pare-Biagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature, 414*, 546-550.
- Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. E. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13*(4), 564-574.
- Posner, M. I., Goldsmith, R., & Welton, K. E. J. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology, 73*(1), 28-38.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*(3), 353-363.
- Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences, 6*(10), 416-421.
- 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., & 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*, 420-428.
- Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998b). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Science, USA, 95*, 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*, 245-257.
- Reed, J. M., Squire, L. R., Patalano, A. L., Smith, E. E., & Jonides, J. (1999). Learning about categories that are defined by object-like stimuli despite impaired declarative memory. *Behavioral Neuroscience, 113*(3), 1-9.
- Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*, 382-407.
- Roediger, H. L. I., Buckner, R. L., & McDermott, K. B. (1999). Components of processing. In J. K. Foster & M. Jelicic (Eds.), *Memory: Systems, Process, or Function?* New York: Oxford University Press.
- Schacter, D. L., & Wagner, A. D. (1999). Medial temporal lobe activations in fMRI and PET: Studies of episodic encoding and retrieval. *Hippocampus, 9*, 7-24.
- Schultz, W., Trembley, L., & Hollerman, J. R. (2003). Changes in behavior-related neuronal activity in the striatum during learning. *TRENDS in Neurosciences, 26*(6), 321-328.
- Scoville, W. B. (1968). Amnesia after bilateral mesial temporal-lobe excision: Introduction to case H.M. *Neuropsychologia, 6*, 211-213.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery and Psychiatry, 20*(1), 11-21.
- Seger, C. A., & Cincotta, C. M. (2002). Striatal activity in concept learning. *Cognitive, Affective, & Behavioral Neuroscience, 2*(2), 149-161.
- 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*, 1316-1324.
- Shafritz, K. M., Gore, J. C., & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Science, USA, 99*(16), 10917-10922.
- Shipp, S. (2004). The brain circuitry of attention. *TRENDS in Cognitive Sciences, 8*(5), 223-230.
- Skudlarski, P., Constable, R. T., & Gore, J. C. (1999). ROC analysis of statistical methods used in functional MRI: Individual subjects. *NeuroImage, 9*, 311-329.
- Smith, D. J., & 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, D. J., Murray, M. J. J., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23(3), 659-680.
- Smith, E. E., & Medin, D. L. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, 7(4), 241-253.
- Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, 65, 167-196.
- Squire, L. R., & Knowlton, B. J. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Science, USA*, 92, 12470-12474.
- Squire, L. R., & Knowlton, B. J. (2000). The medial temporal lobe, the hippocampus, and the memory systems of the brain. In M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (2nd ed.).
- Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, 27.
- Stark, C. E. L., & Okado, Y. (2003). Making memories without trying: Medial temporal lobe activity associated with incidental memory formation during recognition. *The Journal of Neuroscience*, 23, 6748-6753.
- Stark, C. E. L., & Squire, L. R. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Science, USA*, 98(22), 12760-12766.
- Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. Stuttgart: Georg Thieme Verlag.
- Tulving, E., Habib, R., Nyberg, L., Lepage, M., & McIntosh, A. R. (1999). Positron emission tomography correlations in and beyond medial temporal lobes. *Hippocampus*, 9, 71-82.
- Van Essen, D. C., Drury, H. A., Joshi, S., & Miller, M. I. (1998). Functional and structural mapping of human cerebral cortex: Solutions are in the surfaces. *Proceedings of the National Academy of Science, USA*, 95(788-795).
- Van Essen, D. C., Lewis, J. W., Drury, H. A., Hadjikhani, N., Tootell, R. B. H., Bakircioglu, M., et al. (2001). Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Research*, 41, 1359-1378.
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, 277, 376-380.

- Vidnyanszky, Z., & Sohn, W. (2005). Learning to suppress task-irrelevant visual stimuli with attention. *Vision Research*, *45*, 677-685.
- Visscher, K. M., Miezin, F. M., Kelly, J. E., Buckner, R. L., Donaldson, D. I., McAvoy, M. P., et al. (2003). Mixed block/event-related designs separate transient and sustained activity in fMRI. *NeuroImage*, *19*, 1694-1708.
- Vogels, R., Sary, G., Dupont, P., & Orban, G. A. (2002). Human brain regions involved in visual categorization. *NeuroImage*, *16*(2), 401-414.
- 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.
- Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*, 953-956.
- Wallis, J. D., & Miller, E. K. (2003). From rule to response: Neuronal processes in the premotor and prefrontal cortex. *Journal of Neurophysiology*, *90*, 1790-1806.
- Witherspoon, D., & Moscovitch, M. (1989). Stochastic independence between two implicit memory tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 22-30.