NORTHWESTERN UNIVERSITY

Instantiating two learning systems in the brain: cognitive modeling and fMRI analyses of visual category learning

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Neuroscience

By

Emi Marie Nomura

EVANSTON, ILLINOIS

June 2008

ABSTRACT

Instantiating two learning systems in the brain: cognitive modeling and fMRI analyses of visual category learning

Emi Marie Nomura

The set of experiments described here test the hypothesis that the declarative memory system supported by the medial temporal lobe and habit/procedural memory supported by the basal ganglia are recruited when learning novel category representations. The theory guiding specific hypotheses about these neural systems and their operation in category learning are incorporated into a new cognitive model of category learning called PINNACLE (Parallel Interactive Neural Networks Active in Category LEarning). PINNACLE contains two competitive learning systems that simulate a fronto-MTL circuit that supports declarative memory and executive attention and cortico-striatal loops through the posterior caudate that support procedural learning. Chapter 1 reviews the relevant background literature on visual category learning including the hypothesized role of independent memory systems. This background includes recent behavioral, neuroimaging and neuropsychological studies examining the MTL and the caudate of the basal ganglia in learning visual categories either by verbalizeable rules or without awareness. Chapter 2 discusses two fMRI studies that preferentially engage either the MTL or caudate according to the learning strategy employed, demonstrating the capability of both of these systems in visual category learning. Chapter 3 describes the mathematical implementation of PINNACLE and some preliminary modeling results applied to fMRI data. Chapter 4 lays out the theoretical basis of the PINNACLE model, establishing the

framework for the implementation of the theory into a neuro-cognitive model. The fifth chapter discusses the application of PINNACLE in enhancing and extending conventionally accepted fMRI data analysis techniques. The results of PINNACLE-based trial sorting both reinforce the original fMRI dissociation and demonstrate the benefit of combining these two methodologies. Finally, in the last chapter, several potential modifications to the mechanisms of the RB and II systems are suggested along with examples of the application of PINNACLE to other learning behaviors.

DEDICATION

The work here would not have been possible without the unwavering support of my family and friends. My parents have always emphasized the importance of education in my life, and I truly believe that whatever motivation and drive I possess can be almost entirely credited to them. I hope to always live up to the potential that they see in me. My grandfather, Carl Nomura, has also been an inspiration throughout my life, and particularly in graduate school. He reminds me that science is everywhere, whether it's looking at the horizon and calculating the distance to the sun or in the digits of Pi. And of course my little brother and sister, my other thirds, you were there for me through it all.

I truly could not have gotten through the daily ups and downs of graduate school without my friends. They were there through heartbreaks, celebrations, failed experiments and many many practice talks. Sonia, Tess, Melissa, Joy, Katie, Jena, Emily, Rob, Jun. You have all enriched my life in different ways. My girls, where would I be without your words of wisdom and unending patience over so many cups of coffee or dirty martinis. Rob, you remain the funniest person I've ever met. And Jun, you are probably the most modest person I've ever met—you are an inspiration to us all! Finally, I want to dedicate this to Jacob, for believing in me despite myself.

ACKNOWLEDGMENTS

I would like to acknowledge a number of people who have guided me both personally and professionally through the last 5.5 years. First, my advisor Paul Reber has probably played the largest role in my development as a scientist. The most important lesson I've learned from him was how to trust in my own abilities. I'm still working on the love of math and ability to multi-task. I need to thank all of the current and previous Reber lab members. In particular, Antonio Gisbert was a great friend and a source of moral support throughout my graduate career. His technical skills and general knowledge about anything and everything continues to amaze me.

Todd Maddox and his student Dasa Zeithamova were both extremely helpful in experimental design and in the various stages of model development. My first project would not have been possible without Todd's enthusiasm and confidence that it would actually work! Todd still holds the record for the most prompt email responder I've ever met.

I am indebted to the faculty and staff at both Northwestern Memorial Hospital and Evanston Hospital for keeping the scanners up and running and helping me collect the data that went into my thesis. Along similar lines, I need to acknowledge Ted Whalen and Kate Murray who were both integral in keeping ENAC up and running and supplementing my knowledge of AFNI and python. Finally, I'd like to acknowledge Jim Houk and Ken Paller, for their role in the progression of my thesis work. Both have helped me look at my data from a different perspective, sometimes a better perspective.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Overview

Categorization can be defined broadly as the act of responding differently to objects and events in the environment based on their belonging to separate classes or groups. Objects within a category typically share certain behavioral characteristics or physical features. Category learning, then, refers to the cognitive process(es) that extract similar features from a collection of stimuli to accurately categorize novel stimuli. This ability reduces the need for a separate response to each object in the world, making behavior more economical. Given the important role of categorization in everyday life, it is not surprising that there is a long history of study on the underlying cognitive operations that mediate this skill.

Investigations into category learning extend both across and within disciplines, reflecting the multi-faceted complexity of this skill. The first section here briefly outlines the prominent theories of categorization and the accompanying paradigms used to test these theories. In the second section, one cognitive theory of category learning is described that draws primarily from neuropsychological evidence, yet makes a number of specific behavioral predictions. The third section outlines recent evidence from neuroimaging, neuropsychology and experimental psychology that support the multiple systems view of category learning. Finally, existing mathematical and computational models that hypothesize different information processing mechanisms for category learning are discussed.

1.2 Theories and tests of category learning behavior

A number of theories exist that attempt to account for the collection of behaviors grouped together as categorization. In most cases, these theories are put forth in the form of a computational model that is then used to fit existing data. The advantage of this approach is that it allows for an evaluation of different theories in a formalized, mathematical manner. Decision-bound, exemplar and prototype theories represent three of the most successful approaches to explaining categorization behavior in the literature. While described briefly here, there is a long and rich literature devoted to each of these theories.

The process of creating a representation of category structure can be described as partitioning perceptual space and assigning category labels or motor responses to regions that encompass a collection of similar stimuli. One formulation of this process is decision-bound theory (DBT) of category learning first proposed by Ashby and Gott (Ashby and Gott, 1988).

Under DBT, when presented with a to-be-categorized stimulus, subjects determine in what region the stimulus has fallen and produce the associated response. The response regions are divided by a decision boundary. In this approach, learning the categories amounts to identifying the decision-boundary that separates the categories in the perceptual space.
A number of reports support DBT as an effective description of visual category learning (Ashby and Gott, 1988; Ashby and Maddox, 1990, 1992). Typically, the stimuli in these experiments vary on 2 dimensions. For example, in one task, subjects are asked to categorize rectangular stimuli that vary in either the length or the width (Figure 1.1A). In another task, the stimuli are circles of different diameters that have an internal line that varies in orientation (Figure 1.1B). The stimuli can also be perceptually more complex, such as sine wave gratings (Figure 1.1C). All of these examples can come from the same category structure, with the only difference being

the stimulus dimensions. The two-dimensional perceptual space is partitioned into 2 (or more) categories by decision boundaries that can be linear or non-linear. A non-linear boundary requires a more complex representation, but even linear boundaries can vary in the demands placed on the category learner.

Exemplar models of categorization (Medin and Schaffer, 1978; Hintzman, 1986; Nosofsky, 1986) posit that people represent categories by storing individual exemplars of the category in memory. In one of the most well known exemplar models, Nosofsky's generalized context model (GCM), exemplars located near each other in representational space are more similar than those far from each other. Both recognition and categorization of a novel stimulus then depend on its similarity to the representations of previously seen exemplars. While recognition depends on similarity to a specific exemplar, categorization relies on a summed similarity to previously seen exemplars.

A single system that is responsible for both categorization and recognition is difficult to reconcile with the observation of intact category learning in amnesic patients (Knowlton and Squire, 1993; Knowlton et al., 1994; Filoteo et al., 2001a) but a mathematical model that attempts to do this has been described (Nosofsky and Zaki, 1999). If a hypothesis for how the model proposed by Nosofsky & Zaki (Nosofsky and Zaki, 1999) could be instantiated in the brain, it may be possible to distinguish between that type of "single system" model with the two-system models of categorization (Reber et al., 2003).

Prototype models (Posner and Keele, 1968; Reed, 1972; Smith et al., 1997) maintain that a category representation consists of a prototype of the trained exemplars. In this case, a prototype corresponds to the central tendency of the experienced stimuli. Novel stimuli are categorized according to their similarity to this prototype. This differs from the exemplar models in that one maintains separate representations of individual instances of experienced stimuli for recognition and the prototype representation for categorization.

1.3 Category learning paradigms

When studying category learning, it is important to remove the potentially confounding effect of experience that may be present with naturally occurring categories (i.e. tools, faces, animals). Creating artificial categories not only ensures that the participant is forming new representations, but also allows the experimenter to control the underlying category structure. Much recent evidence suggests that the information structure of the category may determine what neural system will support learning (Ashby and Ell, 2001; Maddox et al., 2003). In the following fMRI experiments (Chapter 2), we took advantage of this observation to study the operation of different neural systems in category learning.

There are a number of different types of category learning tasks that are used frequently in the literature. They can be grouped broadly into 4 different types: rule-based tasks, information-integration tasks, prototype distortion tasks and probabilistic learning (weather prediction) tasks. Depending on the particular task demands, people employ different cognitive strategies that rely on the operation of distinct neural systems. Different theories of categorization have relied on these tasks to compare the predictions of the models to human behavior.

1.3.1 Rule-based and information-integration tasks

In DBT, a linear boundary that segments the perceptual space along one dimension (e.g., a horizontal or vertical boundary) creates 2 categories that can be easily described by a verbal rule. In contrast, a linear decision boundary that does not fall along a cardinal orientation requires the learner to integrate information across the 2 dimensions in order to determine category membership. In the first case, the category structure is considered rule-based (RB) in that a simple rule describes the categories. In the second case, determining the category structure requires information-integration (II) and cannot be accomplished using a simple rule. In RB category structures, participants tend to use an explicit reasoning process consisting of one or more verbalizable rules to learn the category (Ashby et al., 1998). Typically, only one of several stimulus features is relevant, so participants can systematically test the different features to discover a rule that will allow for accurate categorization. For example, in Figure 1.2A the optimal decision boundary is a uni-dimensional rule that only depends on the frequency of the stimuli. In II tasks, category membership is best determined by integrating two or more stimulus dimensions before making a category judgment (Ashby et al., 1998). An important characteristic of the II task is that the optimal strategy is very difficult to verbalize. As you can see in Figure 1.2B, accurate categorization can only be achieved by incorporating both frequency and orientation information.

1.3.2 Prototype distortion task

In the prototype distortion task, which has been used largely to test the prototype theory, a prototype category member is defined (traditionally, as a pattern of 9 dots), and then randomly distorted to create category members (Posner and Keele, 1968). During training in this task, participants are instructed to point at the center dot and are not informed that there is an underlying category (although in an explicit version of this task, participants were informed before training to pay attention to the category members; (Reber et al., 2003)). At test, participants endorse new category members at a level greater than chance without necessarily being aware of or even having seen the underlying prototype. Various degrees of distortion from the prototype can create different levels of category members, and participants are sensitive to this variation. If the degree of distortion is high, the exemplar stimuli are less similar to the prototype and with low distortion, the stimuli are more similar to the prototype. Categorization performance is best for the prototype followed by low and then high distortions of the prototype.

1.3.3 Probabilistic learning task

The weather prediction task is one of the most popular probabilistic learning paradigms first used by Knowlton and colleagues (Knowlton et al., 1994) to test the category learning abilities of amnesic patients. On each trial, subjects are shown cards with different combinations of geometric patterns and asked to predict the outcome (rain or sun). Following the subjects' decision, they receive feedback, which allows them to learn the relationships over time. The feedback is probabilistically associated with different cards such that perfect accuracy is not possible. The probabilistic nature of the task is a feature designed to discourage explicit memorization of the cards.

1.4 COVIS: A Neurocognitive theory of category learning

The COVIS theory (COmpetition between Verbal and Implicit Systems) proposed by Ashby (Ashby et al., 1998) describes a neurobiologically inspired instantiation of DBT with specific predictions about the neural systems supporting two separate types of category learning (RB and II). In this theory, 2 learning systems compete to provide the output response: an explicit, rule-based system dependent upon working memory and attention; and an implicit, procedural learning system.

While the COVIS theory is based on the connections and computational properties of cortico-striatal and frontal circuits, the parallels between the multiple neural systems theory of categorization and multiple memory systems of the brain is of note. Studies of memory dating to Scoville and Milner (Scoville and Milner, 1957; Squire, 1992) have established an important difference between conscious, declarative memory based on the medial temporal lobe (MTL) and a collection of heterogeneous nondeclarative memory systems. Studies of nondeclarative memory have shown the importance of the basal ganglia for some nondeclarative memory tasks, including category learning tasks (Knowlton et al., 1996a). Integrating these two conceptual frameworks is a fairly straightforward process that should provide additional organizing principles to the model as a whole. The hypothesized organization of the neural circuits supporting RB and II category learning is shown in Figure 1.3. This schematic is adapted from the COVIS theory of Ashby et al. (Ashby et al., 1998) but emphasizes the role of the MTL in supporting RB category learning as predicted by memory systems research.

1.4.1 RB system in COVIS

Under the COVIS theory, the RB system learns through a conscious process of rule generation and testing, cognitive functions normally subserved by the frontal lobes. Neuroimaging of RB tasks has shown consistent activity in the prefrontal cortex (PFC), anterior cingulate cortex (ACC) and head of the caudate (Rao et al., 1997; Lombardi et al., 1999; Rogers et al., 2000), all areas implicated in working memory and executive attention. COVIS theory assumes that potential rules are stored in working memory while being tested, and depending on the feedback, the rule is either discarded or retained. One prediction from this idea is that the presence of a concurrent task should preferentially impact learning in the RB system as it would interfere with working memory function.

Memory systems research emphasizes the critical role of the MTL is both acquisition and retrieval of conscious, declarative long-term knowledge that contributes to rule-based processing. We hypothesize that the role of the MTL is to store the specific rule and/or referent for the decision boundary that distinguishes the categories. Consistent with this idea, COVIS has been updated recently to include this structure within the RB category learning network (Ashby and Valentin, 2005; Ashby and Ennis, 2006). Because the MTL is hypothesized to be exclusively involved in declarative memory supporting RB categorization, it is a crucial brain region for differentiating between the neural correlates of RB and II category learning.

1.4.2 II system in COVIS

The II category learning system in COVIS is hypothesized to depend upon the posterior body and tail of the caudate nucleus and its interconnections with posterior visual cortical areas. A core element of the COVIS theory is that this system is non-verbal and implicit. From a memory systems perspective, the II category learning element of COVIS describes a specific hypothesis about the structure and function of the type of nondeclarative memory that supports category learning. This system is hypothesized to depend on several important neurobiological properties of the caudate which may make this structure ideal for visual category learning. The many-to-one projections of the visual cortical neurons in TE (inferotemporal cortex) onto the spiny neurons of the tail of the caudate means that thousands of cortical neurons synapse on individual spiny neurons (Wilson, 1995). This massive convergence allows a wide variety of complex information to be reduced (compressed) to its most basic representation, which is precisely the type of process that is necessary for categorization. This lower resolution representation can then be used to categorize novel stimuli.

Another neurobiological property of the caudate that facilitates this type of learning is the existence of cortico-striatal loops that project from specific cortical regions to the basal ganglia and back to these same cortical regions (Middleton and Strick, 2000; Kelly and Strick, 2004; Houk, 2005). Just such a loop from inferotemporal-cortical areas through the posterior portion of the basal ganglia and back (Yeterian and Pandya, 1995) could be particularly important for visual category learning.

The final property of the caudate that is relevant for this implicit category learning system is the availability of a dopamine-based reward signal within the tail of the caudate. Several different forms of evidence indicate that feedback is an important component of this learning process. Animal work has demonstrated that the gluatamatergic projections from visual cortex and the dopaminergic projections from substantia nigra both synapse on the dendritic spines of caudate medium spiny cells (Difiglia et al., 1978; Freund et al., 1984; Smiley et al., 1994; Maddox and Ashby, 2004). Thus, when the system receives an unexpected reward, dopamine is released from the substantia nigra onto medium spiny cells in the caudate and the most recently active synapses responsible for the correct response are strengthened (Wickens, 1990; Schultz, 1992). Given the important role of feedback in this II system, any manipulation of feedback should affect II but not RB learning.

1.5 System interaction

The aforementioned behavioral, neuropsychological and neuroimaging evidence make a strong argument for the functional and anatomical independence of these systems, but more recently investigators have begun to examine the potential forms of interaction between these systems. In the PINNACLE model we assume that the nature of this interaction is competitive, such that the dominant system suppresses the activity of the opposing system.

1.5.1 Competitive learning systems

Evidence from neuropsychology seems to indicate differential and independent roles for the MTL- and caudate-based memory systems. That is, damage to one system does not necessarily interfere with or facilitate the function of the remaining intact system. There is evidence from animal studies as well as human neuroimaging experiments, however, for the existence of a competitive interaction between memory systems. From an evolutionary standpoint, it seems reasonable that multiple learning systems would arise from the variety of different learning demands in the environment. At the most basic level, one requires a system capable of learning specific features and a system capable of learning invariant attributes across events. It would seem that the MTL and striatum are ideally specialized to meet these disparate learning requirements. Competition between the systems should result in a selection of the system that is most suited to learn in a given situation while suppressing the activity of the non-advantageous system.

1.5.1.1 Animal studies

One route to testing this alleged competitive interaction is to systematically inactivate the operation of one of the systems and allow learning to proceed in the other system. Under different learning demands, a facilitation in performance would reveal the conditions in which learning in the intact system is enhanced without the competitive influence of the non-optimal system.

In one of the first studies to find a double dissociation between the behavioral effects of hippocampal and caudate-lesioned animals (Packard et al., 1989), Packard and colleagues found evidence that these systems may be in direct competition. Rats with bilateral lesions of the caudate or fimbria-fornix (impairs hippocampal function) were tested in two varieties of a radial maze designed to favor either hippocampal or caudate-based memory. In the ‘win-shift’ paradigm (Olton and Samuelson, 1976), no single stimulus is consistently paired with the correct response. The rat must learn to only search those arms of the maze that it has not previously visited. Learning in this task has been shown to be hippocampal-dependent. In the ‘win-stay’ task, a single stimulus was consistently paired with the correct response so that acquisition involved learning to approach a specific sensory cue over a number of repeated trials. This task is hypothesized to depend on the operation of the caudate nucleus. As expected, lesions of the fimbria-fornix impaired learning of the win-shift task and lesions of the caudate impaired learning of the win-stay task. Interestingly, when compared to control animals, there was an observed facilitation of learning in the fimbria-fornix lesioned animals on the win-stay task. The facilitative effect of fornix lesions in this study suggests that the hippocampal and caudate systems were in competition and when the hippocampal system was removed, the caudate system was able to learn more efficiently. In a similar study where lesions were made after training, inactivation of the hippocampal system was still shown to enhance caudate-based response learning (Schroeder et al., 2002). Thus, the competitive interference between these systems may occur both during task performance as well as during the consolidation period.

In addition to demonstrating the competitive influence of one memory system on another, researchers have observed a natural shift between system operation in the untreated rat. Packard (Packard and McGaugh, 1996) trained rats for 14 days to approach a consistently baited arm in a cross-maze starting from the same initial position. Their memory was probed on the 8th and the 16th day by placing them in a different starting box and observing which arm of the maze they entered. If they entered the same maze arm that had previously been baited, they were designated as 'place learners'. If they entered the unbaited maze arm, which meant recalling the series of learned motor patterns (i.e. always turning to the left), they were designated as 'response learners'. Rats that had been treated with saline displayed place learning on the first probe trial and response learning on the second probe trial, which indicates a natural shift in learning mechanism to guide behavior. That is, early in learning, rats relied on hippocampal-based mechanisms to recall the specific location of the food, whereas later in learning they seemed to rely on a more automatic, habit-like learning supported by the caudate-based system. Rats that were given a lidocaine injection (a local anaesthetic that produces neural inactivation) into the hippocampus showed no preference for place or response learning on the first probe trial, but displayed response learning on the second probe trial. The time course of this shift was sensitive to manipulation by preferentially enhancing one or the other system through infusions of glutamate into either the hippocampus or caudate (Packard, 1999). Glutamate is an excitatory neurotransmitter that in previous studies has been shown to enhance memory (Packard and Teather, 1999). Hippocampal glutamate injections strengthened the place learning representation, effectively blocking the shift to response learning in the second probe trial and caudate injections accelerated the shift to response learning such that it was detected in the first probe trial. Post-training injections of glutamate were thus shown to modulate the relative engagement of the hippocampus and caudate by enhancing the normal shift from one system to another.

While we cannot necessarily relate maze-learning in rats directly to a more complex skill such as category learning in humans, an understanding of the basic operation of these memory systems is essential to developing theories and models of memory. Based on the animal data here, there is some evidence that in certain learning environments, hippocampal- and caudate-based memory systems compete to provide the output response.

1.5.1.2 Neuroimaging studies

Neuroimaging of a number of different skill learning paradigms support a competitive interaction between the MTL and caudate-based memory systems. This data suggests a general pattern of interaction across learning paradigms, which bolsters our confidence that PINNACLE is implementing the appropriate neural mechanisms for interaction between these systems in category learning.

In one of the first neuroimaging studies to put forth evidence for competitive MTL and caudate-based memory systems, Poldrack and colleagues (Poldrack et al., 1999) demonstrated that the caudate activated and the MTL deactivated during a weather prediction task. The involvement of the caudate was expected, based on the poor performance on this task by patients with Parkinson’s and Huntington’s disease (Knowlton et al., 1996a; Knowlton et al., 1996b). The observed deactivation in the MTL, however, was not expected and an analysis of the time course of activity showed that the negative signal change became more pronounced up to a point, and then disappeared later in learning. The authors interpreted this decrease in deactivation as the MTL coming on-line towards the end of the task, which was consistent with the observation that amnesic patients are impaired on this task late in training (Knowlton et al., 1994). When this activity was compared to that during another version of the task that relied on observational learning, the MTL deactivation was shown to be specific to the original feedback-based paradigm. Further, a correlational analysis using the task-evoked signal in the MTL determined that activity in the caudate nucleus was negatively correlated with the MTL signal (Poldrack et al., 2001). When Parkinson's patients were scanned while learning this same task, they demonstrated more activity in the hippocampus and less activity in the basal ganglia than control subjects (Moody et al., 2004) suggesting that a lack of competitive suppression allowed the explicit memory system to participate when it normally would not. Based on this result, the authors hypothesized that healthy participants may utilize an explicit strategy early on, but eventually abandon it (or it is competitively suppressed) as the striatal system proves to be more successful at learning the probabilistic feedback contingencies. A contrary interpretation of the PD finding is that in disease, the functional system will always take over, so it was not that the explicit system was released from competitive suppression, but that it was simply the only operational system in the brain for the task.

The apparent competitive interaction between the MTL and basal ganglia has also been observed in a number of other neuroimaging studies that did not use the probabilistic category learning task. In a task that focused more on rule learning with no probabilistic component, fMRI revealed a similar pattern of opposing activity in the caudate and MTL (Seger and Cincotta, 2006). Participants were asked to acquire a particular rule that required attending to different features of the letter stimuli. Using the feedback after each trial, they eventually discovered the rule and then applied it to the subsequent stimuli. Rule learning relied upon activity in a wide network of frontostriatal areas coupled with a decline in hippocampal activity whereas in rule application, there was a rise in hippocampal activity. The authors suggest that the antagonism between the striatum and MTL may have facilitated the transition between acquisition and application of a rule.

A recent fMRI study supporting a competitive form of interaction between declarative and habit learning systems found that the presence of a distracting task modulated the degree to which the subject engaged the striatum (habit learning) or the hippocampus (declarative memory) (Foerde et al., 2006). Furthermore, the resulting information that was available after learning depended on the conditions under which it was acquired. When learning in the presence of distraction, later expression of knowledge was less flexible and reliant on the information acquired by the striatal system. Under non-distracting conditions, both the MTL and striatal systems were active, yet later knowledge was more flexible.

In both motor skill learning (Jenkins et al., 1994) and perceptual skill learning (Poldrack and Gabrieli, 2001), increasing activation was observed in the striatum accompanied by a deactivation in the MTL. We also observed a similar pattern of activity within our own data using a visual category learning paradigm (Nomura et al., 2007). Taken together, these separate observations seem consistent with a competitive interaction between these memory systems. However, there is also an important caveat associated with observed deactivations in neuroimaging studies. Little is known about the neurophysiological mechanism behind these deactivations, and it is not clear if inhibitory synaptic activity is associated with an increase or a decrease of fMRI signal.

1.5.2 Other modes of system interaction

1.5.2.1 Parallel independent learning systems

Despite the numerous forms of evidence in favor of competitive MTL- and caudate-based memory systems, there is also evidence that these systems may exist in relative isolation, operating in parallel as they each contribute toward learning a particular task that is suited to that system. In this framework, learning in one system may not necessarily help or hurt learning in the other system. If both systems are simultaneously acquiring information about the categories, damage or interference with one system should not affect the representation in the other system.

Perhaps the most conspicuous evidence for independent learning systems comes from neuropsychological double dissociations of MTL and caudate-based memory. For example, amnesic patients are impaired at recognition memory but exhibit normal motor skill learning (Tranel et al., 1994) and Parkinson's patients show the opposite pattern of impairment (Heindel et al., 1989). Similarly, in rats that have received reversible lesions to the hippocampus or caudate, learning progresses normally, presumably relying on the operation of the preserved system (Packard and McGaugh, 1996; White and McDonald, 2002). Using a radial arm maze task in which learning can be supported by either system, pre-training lesions to the hippocampal or the caudate system did not affect the rat's ability to acquire the task (McDonald and White, 1994). Only when simultaneous lesions were made to both systems was learning impaired.

A similar observation was made in humans combining neuropsycholgy and neuroimaging (Voermans et al., 2004). Patients with Huntington's disease were scanned during a route recognition paradigm and found to rely on the hippocampus to maintain normal behavior. This stands in contrast to healthy subjects who use both the MTL and caudate during this task. This suggests that in a healthy brain, both systems play a role in acquiring different sorts of information, the hippocampal system may create a map-like representation and the caudate system a more route-like representation. Either system may be capable of supporting accurate performance, and in the situation where one representation is missing (i.e. in Huntington's disease), the other system can compensate. This compensation may not be bidirectional, though, and further studies are needed to address this question. From this study, it seems that hippocampal-based memory can supplement caudate-based memory, but there is little evidence for the opposite phenomenon.

While evidence suggests that both the hippocampal- and caudate-based systems can be active simultaneously, the system that is most relevant to a given task ultimately comes to dominate behavior. This does not necessarily imply competitive systems, as the information in the non-dominant system may still be accessible. In the same cross-maze task discussed earlier (Packard and McGaugh, 1996), rats that were given a lidocaine injection (a local anaesthetic that produces neural inactivation) into the hippocampus showed no preference for place or response learning on the first probe trial, but displayed response learning on the second probe trial. In contrast, rats that were given a lidocaine injection into the caudate displayed place learning on both probe trials. Together, these results indicate that in a task where these two systems are each capable of supporting learning, a neurochemical inactivation of one system does not interfere with the ability of the remaining system to acquire the information.

1.5.2.2 Multiple modes of interaction

Another potential scenario of system interaction is that early in learning, the systems participate in one form of interaction, and at later stages of learning or perhaps at the point of expertise, this interaction changes. For example, it could be the case that the systems compete to learn the task in initial stages of acquisition, and as it becomes apparent that one system is better suited to learn, there is no longer a need for this competition.

It is also possible, as suggested by Mcdonald and White (McDonald and White, 1995), that certain learning situations engage different types of interactions amongst the different memory systems. In the context of rat maze-learning, when the information processing demands are such that either system could acquire the representation necessary to produce accurate performance, the systems would be acting cooperatively. In contrast, in cases where the information acquired by one system interferes with the accurate performance guided by the other system, competition between the systems needs to be resolved. This will often result in the need for a larger number of training trials (McDonald and White, 1993; White and McDonald, 1993) to accurately perform the task.

1.5.3 Mediators between the RB and II systems

There are a number of different potential mechanisms for system mediation between the RB and II systems based mainly on animal research. The first is through direct anatomical connections between the MTL and striatum. In rats, there is evidence for direct projections from the entorhinal cortex to the dorsal striatum (Sorensen and Witter, 1983) that are inhibitory in nature (Finch, 1996). Similar projections from the striatum to the MTL have been difficult to demonstrate in primates, but caudate stimulation in cats induces a theta rhythm in the hippocampus (La Grutta and Sabatino, 1988) effectively disrupting natural spiking activity in that area. This dampening of activity in the hippocampus may be mediated by cholinergic activity as administration of an anti-cholinergic agent attenuates the effect. This could be one potential mechanism for the frequent deactivation observed in the MTL in functional imaging data, although the anatomical connections need to be established.

A more likely possibility is that a separate region is mediating the interaction between the MTL and striatum. The best candidate for the mediating factor between the hippocampal- and caudate-based memory systems is the PFC, based on its known role in executive control functions. Under this hypothesis, learning progresses independently in the MTL and striatal-based memory systems, but the PFC's role is at the level of response selection. Evidence for the mechanism of this hypothesized interaction comes again from experiments with rats. Hippocampal gluatamatergic projections have been shown to synapse on PFC neurons, forming a transcortical network that mediates cognitive processes such as working memory (Gabbott et al., 2002). When these afferents are stimulated, it can evoke both excitatory and inhibitory responses in PFC neurons (Lewis and O'Donnell, 2000). In rats, there exist direct inputs from MTL and striatum that converge on common PFC regions, suggesting a possible site of interaction of these systems (Floresco and Grace, 2003). When the mediodorsal (MD) thalamus (output region of the basal ganglia) was stimulated while the hippocampal-PFC circuit was concurrently stimulated, the PFC firing was inhibited. Similarly, when hippocampal inputs were stimulated, the PFC response to MD stimulation was inhibited. It remains to be determined if the neurons within the MD thalamus received striatal input in this particular study, although this nucleus is included in known limbic corticostriatal loops (Joel and Weiner, 2000).

While we are developing a model with specific assumptions of interactive mechanisms to address the question of system interaction in humans, analyses of fMRI data that attempt to assess the connectivity between neural structures provide an alternative method to test the same idea. Using the fMRI data discussed previously (Poldrack et al., 2001), the authors conducted a path analysis based on the time series data extracted from eight brain regions active in the task (Poldrack and Rodriguez, 2004). The idea behind this method is to calculate the path coefficients between regions to determine the index of causal influence between those variables. In addition to the MTL and striatum (dorsal and ventral), they also examined the connectivity between the ACC, dorsolateral PFC (DLPFC), midbrain (substantia nigra/ventral tegmnetal area), fusiform gyrus and superior parietal cortex. When they restricted the analysis to only those connections that were neuroanatomically plausible, the results argued strongly for prefrontal mediation of memory system interactions and against direct interactions between the striatum and MTL. Specifically, Poldrack (Poldrack and Rodriguez, 2004) observed negative reciprocal paths between MTL and PFC and positive paths between the striatum and both the MTL and PFC.

While there exists evidence for a number of different potential mediating mechanisms between the RB and II systems, in PINNACLE we assume it is the PFC that resolves the competition between systems. As discussed in Chapter 5, these assumptions can be tested formally by combining the model predictions with fMRI data. If the model that includes this PFC mechanism for resolving competition between the systems is used to examine the fMRI data, we can test whether or not activity in the PFC is actually associated with instances of model-identified competitive interaction.

1.6 Evidence for independent systems supporting category learning

1.6.1 Neuropsychological studies of category learning

Category learning was originally shown to be dissociable from declarative memory in amnesic patients who exhibited normal learning of dot-pattern categories (Knowlton and Squire, 1993), artificial grammar (Knowlton et al., 1992), and a probabilistic classification task (Knowlton et al., 1994; Reber et al., 1996). In each case, patients with severe impairments in declarative memory for previously seen stimuli exhibited unimpaired category learning. It is important to note that although Knowlton and colleagues (Knowlton et al., 1994) observed normal learning by amnesic patients in the probabilistic classification task, in the later trials, the participants were impaired relative to controls. In another experiment using this same task, amnesics with focal bilateral hippocampal damage were impaired throughout the experiment (Hopkins et al., 2004), suggesting a specific role of the hippocampus in this task. However, across these studies of amnesic patients, it appears that nondeclarative memory in some cases can contribute toward category acquisition when declarative memory is impaired.

The focus on the caudate in the COVIS model has meant that the majority of studies using RB and II tasks have examined the performance of patients with dysfunction of the basal ganglia (e.g., Parkinson's and Huntington's disease; PD and HD respectively). While the typical symptoms of these syndromes are motor-related, cognitive deficits are also observed, one being category learning. There is, however, some disagreement in the literature as to the specific category learning deficit in PD patients. PD patient performance on a perceptual categorization task with simple line stimuli varying in horizontal and vertical length was impaired only with non-linear II decision boundaries but was no different from controls with a linear RB decision bound (Maddox and Filoteo, 2001). Similarly, HD patients showed a selective impairment in II categorization (Filoteo et al., 2001b).

Conversely, in several additional studies, PD patients showed impairments in RB learning using geometric figures (Ashby et al., 2003b) and with stimuli of varying line lengths (Maddox et al., 2005a; Ell et al., 2006), but showed no impairment in II category learning.

These conflicting results can be resolved by considering the difficulty of the tasks. The simple line stimuli used in the II category learning task may not have been difficult enough to challenge the patients. That is, patients could have used a sub-optimal RB strategy in the II condition, perhaps relying on the intact operation of a frontal/MTL network. In a follow-up study, Filoteo and colleagues (Filoteo et al., 2005) examined II category learning in greater detail in PD patients using linear and non-linear II category structures. The complexity of the II task affected PD patients in that they were impaired relative to controls in the non-linear version of the task. Computational modeling of this data identified participants who used an II approach, and out of these, PD patients showed a specific impairment in non-linear category learning compared to controls as observed previously (Maddox and Filoteo, 2001).

The question of the RB category learning deficit that is present in some (Ashby et al., 2003b; Maddox et al., 2005a; Ell et al., 2006) but absent in other studies (Maddox and Filoteo, 2001) can be resolved by taking into account the number of irrelevant dimensions present in the task. A greater number of irrelevant dimensions require more selective attention to direct focus to the relevant dimension(s). Those studies that observed a deficit in PD patients' ability to learn RB categories all had a higher number of irrelevant stimulus dimensions than those that did not observe a deficit. In a recent study (Filoteo et al., in press), this issue was addressed by systematically manipulating the selective attention requirements during RB category learning. As predicted, PD patients were found to be impaired at learning RB categories when selective attention demands were greatest. One well documented impairment in PD patients is in their ability to switch between relevant dimensions or rules, such as in the Wisconsin Card Sorting Task (Brown and Marsden, 1988). Thus, the RB deficits seen in these studies may be due to the patients' inability to switch between rules and not necessarily in rule acquisition.

From the aforementioned patient data we can conclude that the intact basal ganglia is required for non-linear, II category learning under low selective attention demands. With increasing task complexity, the dissociation between RB and II category learning becomes less evident. The RB learning deficit seen in these patients may have been due to their difficulty with selective attention that was not relevant in the more simple tasks.

Given the hypothesized role for the MTL in RB category learning, one would expect that damage to the MTL in amnesic patients should also produce impairments on tasks that rely on an RB strategy. However, several studies have reported that amnesic patients perform normally in simple RB category learning (Leng and Parkin, 1988; Janowsky et al., 1989). This suggests that when the rule is simple enough to be held in working memory (intact in amnesia), the damage in the MTL does not affect their ability to categorize. With a more complex category structure, amnesics are impaired at retaining verbalizeable rules (Kitchener and Squire, 2000) although they can learn some complex categorization tasks implicitly (Reed et al., 1999). From the perspective of the dual-system COVIS model, it may be that the intact II system is capable in some cases of supporting learning when the MTL is damaged. It follows then that amnesic patients could perform normally on II category learning as long as explicit memorization is not an alternative strategy that gives control subjects an undue advantage (Filoteo et al., 2001a).

The complexity of category learning and the involvement of the basal ganglia in both II category learning and RB category learning (via dopaminergic support of the PFC and working memory) means that clear double dissociations in the neuropsychological literature have not been observed with II and RB tasks. However, the overall pattern of impairments across patient groups does suggest that distinct neural systems are involved in RB and II category learning.

1.6.2 Review of RB and II behavioral experiments

A series of behavioral experiments have demonstrated dissociations between RB and II category learning systems. Based on the hypothesized properties of the RB and II systems in the COVIS theory, these experiments focus on 4 main manipulations: the presence of a concurrent task, the nature and timing of the feedback, mapping of the response keys and discontinuous categories. Behavioral differences created by these manipulations suggest that there are fundamental differences in the operating characteristics of RB and II category learning.

The RB system in COVIS is hypothesized to depend on maintaining rules (and/or boundary reference information) in working memory during learning. The concurrent presence of a working memory task should thus preferentially impair RB but not II learning. This prediction was tested with a numerical analog of the Stroop task (Waldron and Ashby, 2001) and with a sequential memory scanning task (Maddox et al., 2004c). Both studies found that RB learning was disrupted whereas II learning was insensitive to a working memory load. A criticism of these studies is again that the number of relevant stimulus dimensions differed between the RB and II conditions. However, even when the number of relevant dimensions was held constant, the working memory task still disrupted RB more than II learning (Zeithamova and Maddox, 2006). In an indirect manipulation of working memory, Maddox and colleagues (Maddox et al., 2004b) demonstrated that RB but not II learning was impaired when learning four (rather than two) categories. Together, these behavioral manipulations of working memory strongly argue for its important role in RB but not II learning.

Another prediction from COVIS based on the important role of working memory and executive attention in the RB system is that this should allow for the possibility of a delayed feedback signal (since the relevant category information is maintained over time in working memory). This is not a property shared by the procedural learning system that is hypothesized to support II learning. The nature and timing of the feedback should be more critical for the II learning system to allow for dopaminergic signals to support learning in the cortico-striatal loop through the tail of the caudate. In one of the first of a series of experiments examining the role of feedback in RB and II learning, Ashby and colleagues (Ashby and Waldron, 1999) found that the mere presence or absence of feedback differentially affects learning. When subjects were asked to learn II category structures without feedback, they typically used unidimensional rules (Ashby et al., 1999) whereas when given feedback, they were able to learn unidimensional and complex decision bounds that lack a simple verbal description (Ashby and Maddox, 1992). This suggests that in situations where II learning is necessary, feedback is required for learning to progress, whereas in the absence of feedback, the default strategy is to utilize a verbalizable RB strategy.

Because the integration of feedback in II category learning is thought to depend on the availability of dopamine signals within the caudate, the ability to use feedback should be constrained by the temporal availability of this signal. A delay in feedback following the presentation of the stimuli, would selectively impair II learning because of the inability to maintain information across the delay via working or long-term memory (which are exclusively involved in RB category learning). Maddox and colleagues manipulated the delay between stimulus response and visual feedback in both RB and II subject groups with both two (Maddox et al., 2003) and four categories (Maddox and Ing, 2005). Using two categories of sine wave grating stimuli, they found that II but not RB learning was disrupted when feedback was delayed by 2.5 seconds or more. This finding was replicated in an experiment that manipulated the feedback delay with four equally discriminable categories (Maddox and Ing, 2005). With four categories, the number of relevant stimulus dimensions and requisite attentional demands were equal between groups (as opposed to two categories where RB learning only required attending to one relevant dimension). The timing of feedback is thus an important quality of the II category learning system irrespective of the number of categories being learned, further supporting the hypothesis that II learning depends on a procedural learning-based system that requires a tight temporal correlation between the motor response and the feedback. In contrast, RB learning was demonstrated to be insensitive to manipulations in feedback timing, supporting the hypothesis that the RB system relies on the ability to sustain information over longer times via working or long-term memory.

If a procedural learning-based system is responsible for the acquisition of II category structures, the same characteristics observed in other types of procedural learning tasks should apply to II learning. Motor learning tasks are commonly used to demonstrate procedural learning where a particular motor pattern is associated with external stimuli (Willingham, 1998). One characteristic of procedural learning that was demonstrated in the serial reaction time (SRT) task, is that changing the location of the response keys interferes with learning but changing the sequence of finger movements does not (Willingham et al., 2000).

Applying this observation to category learning, Ashby and colleagues (Ashby et al., 2003a) found that RB and II category learning were differentially affected by a change in the location of the response keys. In this experiment, subjects learned either RB or II category structures under 3 different conditions. In the control condition, they pressed the ‘A’ button with their left hand and the ‘B’ button with their right hand. In the hand-switch condition, their hands were crossed during training so the button assigned to category A was pressed with the right hand and category B with the left hand. During the transfer test, they uncrossed their hands. In the button-switch condition, training was the same as the control condition except at transfer, the button assignments were remapped such that the left hand was now pressing category B and the right category A. RB category learning was found to be insensitive to hand- and button-switch manipulations, but II category learning was disrupted only in the condition where the buttons were switched at transfer. That is, the manipulation that required a re-mapping of the response location selectively disrupted II category learning.

In an extension of this study (Maddox et al., 2004a), subjects were asked to learn categories 'A' and 'B' and make responses in one of 2 configurations. In the consistent stimulus-response mapping group, they responded with button A in the left hand and button B in the right hand. In the variable stimulus-response mapping group, they were asked to respond to the question "Is this an A?" or "Is this a B?" with a 'yes' button in the left hand or a 'no' button in the right hand which meant that the categories required inconsistent motor responses to learn. They found that changing the response keys interfered with II not RB performance. The similarity of these results to Willingham's (Willingham et al., 2000) SRT results supports the claim that II category learning relies on a procedural learning-based system that is highly sensitive to the mapping between the stimulus and response position.

Together, the data from motor-response manipulations suggests that II category learning is mediated by a procedural-learning based system that learns to associate a category label with a response location. This stands in contrast to the RB category learning system that is hypothesized to acquire abstract category labels that are insensitive to motor-response manipulations.

The fourth dissociation between RB and II category learning predicted by COVIS was observed by manipulating the distribution of stimuli within the categories (Maddox et al., 2005b). Studies in both humans and animals indicate that procedural learning systems are dependent upon stimulus similarity and coherence (Cohen et al., 1997), so discontinuous clusters within a category should disrupt II but not RB categorization. RB category learning should be insensitive to experimentally imposed discontinuities in the category structure as learning depends on more abstract verbalizable rules—the rule could be applied regardless of the participant's experience with the particular region in stimulus space. When the continuity of the categories was manipulated (Maddox and Ing, 2005), assigning discontinuous clusters of stimuli to the same category label led to poor II category learning but had no effect on RB learning. To address the potential confound of difficulty due to mismatched dimensions, a follow-up study (Maddox et al., 2007) found that with discontinuous RB and II category structures with equivalent numbers of relevant dimensions, II and not RB learning was still impaired.

As shown here, the COVIS theory successfully predicted a number of behavioral dissociations between RB and II categorization. While these dissociations demonstrate that the operating characteristics of RB and II category learning differ such that they are differentially sensitive to a variety of manipulations, they cannot directly address the question of neural structures engaged nor neurobiological mechanisms responsible for learning. Experimental techniques from cognitive neuroscience, i.e., neuropsychology and neuroimaging, are necessary to bridge from empirical cognitive psychology to neurobiology.

1.6.3 Neuroimaging category learning

Collecting functional neuroimaging data while subjects are undergoing category learning allows us to test hypotheses about the proposed involvement of various cognitive processes in theoretical models of categorization. Recently, a number of neuroimaging studies have investigated the neural correlates of category learning using a variety of different tasks. In general, the data suggest that there exist multiple neural systems that can support category learning, but the system engaged depends on the task demands. The areas that have been identified consistently across these different tasks (basal ganglia, MTL and extrastriate visual cortex) are also those implicated in cognitive models of categorization such as COVIS.

Neuroimaging of the weather prediction task has consistently demonstrated the involvement of the basal ganglia (caudate nucleus) in category learning (Rao et al., 1997; Rauch et al., 1997; Lombardi et al., 1999; Poldrack et al., 1999; Rogers et al., 2000; Seger and Cincotta, 2005). The MTL has also shown to be involved in probabilistic learning, with activity present early in the training session followed by deactivation as training progresses (Poldrack et al., 1999; Poldrack et al., 2001). The authors interpreted this pattern of opposing activity in the MTL and caudate late in training as some of the first evidence for competition between memory systems. That is, early on in the task, the MTL is perhaps mediating explicit memorization, but later in learning, the MTL activity is suppressed as the caudate-based system comes online. When Parkinson's patients were scanned while learning this same task, they showed activation in the hippocampus and less activity in the basal ganglia than control subjects (Moody et al., 2004) suggesting that a lack of competitive suppression allowed the explicit memory system to participate when it normally would not. Based on this idea, healthy participants may utilize an explicit strategy early on, but eventually abandon it as the striatal system proves to be more successful at learning the probabilistic feedback contingencies. A challenge to this idea, however, comes from the finding of late learning deficits in amnesic patients performing the weather prediction task (Knowlton et al., 1994). Healthy subjects maintained an advantage over amnesics late in learning (after 50 trials) presumably by utilizing an additional MTL-based explicit memory strategy.

In a neuroimaging study of the dot pattern prototype distortion task, a dissociation between learning systems was observed in the MTL and visual association cortex (Reber et al., 2003) depending on whether categories were acquired explicitly or implicitly. Subjects were either told that the patterns were all members of the same category before the experiment began (explicit group) or were asked to simply observe the patterns and were told later that they came from the same category (implicit group). Both groups received a subsequent categorization test in which they saw category members and non-members. Scanning during this post-test revealed group differences in activity in the hippocampus and posterior visual cortex. When participants acquired the dot pattern categories explicitly, activity was observed in the left hippocampus greater for categorical than non-categorical trials (Reber et al., 2003). In contrast, decreased activity in the right posterior occipital cortex for categorical more than non-categorical trials suggested that the categorical patterns were processed more fluently.

Finally, in a recent fMRI study of RB and II visual category learning (see Chapter 2), we observed the preferential recruitment of the MTL during RB categorization and the posterior caudate during II categorization (Nomura et al., 2007). These groups were matched in difficulty, contained stimuli that were highly similar, yet participants learned to discriminate between categories successfully over experience with 320 trials. When the appropriate learning strategy was successfully applied, the MTL and the caudate were differentially engaged in RB and II learning, respectively.

Neuroimaging of a variety of categorization paradigms strengthens the claim that there are distinct neural structures capable of supporting category learning depending on the particular type of learning required. A declarative, RB neural network, which includes at least the MTL, can support explicit learning of associations or memorization of categorical patterns. There also exist at least 2 different types of implicit memory that rely on either the caudate or the posterior occipital cortex depending on whether the learning is incremental and feedback-driven or involves some form of perceptual fluency.

1.7 Computational models of categorization

A number of computational models exist that have successfully captured human category learning behavior. While these models often are able to predict the same behavior, the basic tenets differ considerably from model to model, particularly in the degree to which the model's mechanisms can be plausibly mapped onto specific neural structures in the brain. It is often the case, such as in exemplar or prototype models, that no claims were made about the neural plausibility at the outset, so it is not always fair to compare these models to ones such as the striatal pattern classifier (Ashby and Waldron, 1999) that was purposefully instantiating a neurobiologically plausible mechanism. However, by comparing the operating characteristics of these disparate models, one can isolate the particular contributions of each model to our understanding of the cognitive processes that support categorization behavior.

1.7.1 Connectionist implementation of COVIS

While the previous section described the theory underlying COVIS, there is also a computational version of the theory that has been used to fit the existing data (Ashby et al., 1998). The model is composed of a network of connected structures (see Figure 1.3) that is constructed in a traditional connectionist architecture. It contains separate verbal and implicit systems that compete on each trial to acquire categorical information. The implicit system implements a decision bound model of categorization similar to the DBT models described previously. The verbal system consists of a rule selection mechanism containing a weighted value that emphasizes more salient rules. Rule switching is mediated by a different mechanism that assesses how probable the selected rule is versus the alternative rules. Perceptual noise is included in the model by adding a random number to the stimulus representation on every trial. Feedback in this model returns to both systems and when a system is incorrect, the network weights are adjusted accordingly. There are additional parameters of momentum and learning rate annealing that are incorporated into the learning rules of the systems. The basic idea is that these parameters reduce large changes in the system's representations after negative feedback.

The COVIS computational model successfully fit a large set of experimental behavioral data and made predictions that fit a number of neuropsychological disorders. However, the large number of included parameters makes it difficult to assess the neurobiological plausibility of the model. In the later discussion of the newly developed model (PINNACLE), we take a similar approach as this COVIS model, but limit the number of parameters to only those that can be directly related to particular psychological constructs.

1.7.2 SPEED model of automaticity

Ashby and colleagues (Ashby et al., 2007) recently described a new model of categorization that accounts for the result of overtraining in information-integration category learning. SPEED (Subcortical Pathways Enable Expertise Development) is considered to be a model of expertise as it captures the automaticity of performance that occurs over extensive experience. This biologically detailed computational model is intended to be an extension of the procedural learning system within COVIS. The theory is that in initial stages of learning an II category structure, one utilizes the basal ganglia-based procedural learning system, but gradually with much experience, control is passed to cortico-cortical connections between sensory association areas and premotor cortex.

Quantitative predictions of SPEED were examined in several different types of categorization experiments. In a single-cell recording experiment in non-human primates, SPEED accurately predicted the pattern of firing within the premotor cortex after extensive practice on a tactile categorization paradigm. SPEED also accurately predicted the decline in striatal cell firing rate in rats that had received extended training in pressing a lever after hearing a tone. Finally, in a simulation of human category learning, SPEED accurately learned two categories as accurately and with the same decrease in RT as human subjects.

1.7.3 Striatal Pattern Classifier

There exist a number of computational models of categorization that have been applied specifically to RB and II category learning. As previously mentioned, Ashby's COVIS theory makes strong behavioral predictions for RB and II category learning based on the properties of the hypothesized learning systems. Another model associated with COVIS that attempts to better account for the implicit/procedural learning system is the striatal pattern classifier (SPC) proposed by Ashby and Waldron (Ashby and Waldron, 1999). In this model, learning depends on associating response labels with different regions of perceptual space. The response region boundaries are not learned directly (as in COVIS), but rather emerge as a consequence of labeling space. The SPC model contains a variable number of striatal units that are responsible for learning a low-resolution map of perceptual space. Through experience with the stimuli (and feedback), each unit becomes associated with one of the category labels. These striatal units are meant to represent the role of the striatum in learning to associate a category label (and associated motor plan) with a group of cells in the visual cortex. While the ‘striatal units’ are not necessarily meant to emulate striatal medium spiny cells, they are intended to be a hypothetical substrate that simulates the neural architecture of the caudate for the purpose of computational modeling. The SPC model has had some success in capturing the subject data from several categorization paradigms.

In a recent study (Maddox et al., 2007), different versions of the SPC model were fit to the data from a four-category learning task, where the number of striatal units used to represent a single category varied across model versions. The number of free parameters assigned to each version of the SPC model increased with increasing numbers of striatal units. That is, the SPC-1 model, which assumed 1 striatal unit for each category, had 5 free parameters that determined the location of the units and 1 free parameter that represented the noise associated with the placement of the units. Similarly, the SPC-2 model (2 units per category) had 13 parameters, the SPC-4 had 29 parameters, and so on. In modeling, the greater the number of free parameters the better the model is at accounting for the variability in the data, but the cost is that there is less of a chance that each individual parameter has some cognitive reality. Comparing fit values across models requires a calculation that takes into account the number of parameters (e.g. Akaike’s Information Criterion).

Maddox and colleagues (Maddox et al., 2007) applied these SPC models as well as a hypothesis-testing model to data where subjects were learning discontinuous categories. In these category structures, distinct clusters of perceptually dissimilar stimuli must be associated with a single category label, and were thus hypothesized to disrupt II learning. Not only were subjects able to learn these discontinuous II category structures (albeit at a lower accuracy rate than the

RB discontinuous categories), but the model fits revealed a bias toward the SPC models rather than the RB model. Furthermore, the participants with the most accurate performance also demonstrated better fits with the SPC models with more striatal units. The authors interpreted this as suggesting that the improvements in some subjects' behavior correspond to a recruitment of more cells and thus provide greater information processing capabilities. Importantly, the use of a neurobiologically-inspired model here highlights the ability of cognitive computational models not only to better characterize behavior, but also to lend some insight into the processing characteristics of the system that is being modeled.

1.7.4 SUSTAIN

The SUSTAIN (Supervised and Unsupervised Stratified Adaptive Incremental Network) model (Love et al., 2004) is designed to capture human category learning behavior by recruiting clusters in response to surprising/new events. Newly recruited clusters are then available to apply to future events and can eventually become prototypes or rules that define the categories. A major strength of SUSTAIN is that it can flexibly acquire a wide variety of behavioral data across distinct paradigms.

There are five basic principles of SUSTAIN. The first is that SUSTAIN is initially directed toward simple solutions. Given a subset of potential outcomes, it selects the most parsimonious option. Practically, this means it begins with a single cluster and adds additional clusters as more complexity is required. The second principle is that perceptually similar category members tend to be grouped together. This similarity-based grouping is advantageous because it removes the need for separate traces to be stored for each individual item. This aspect of SUSTAIN is comparable to the formation of a prototype that reflects the central tendency of a group of similar items. The third principle in SUSTAIN is that it is able to learn in the presence or absence of feedback. When feedback is present and indicates an item is incorrectly grouped with an existing cluster, a new cluster is formed. In the case of unsupervised learning, new clusters are recruited whenever the similarity of the item is too distant from any of the existing clusters. The fourth principle of SUSTAIN is that the pattern of feedback affects cluster recruitment. As mentioned previously, when the model receives disconfirming feedback, a new cluster is formed, so depending on the nature of the feedback the model will develop differing representations. The final principle is that clusters are in competition with each other. If several clusters have been created that are highly similar to each other, the strength of the dominant cluster will be weakened if the input stimulus is more similar to the non-dominant clusters.

SUSTAIN shares some conceptual characteristics with the SPC model. While the SPC model uses the term ‘striatal unit’, the idea of labeling regions of stimulus space according to experience is similar to creating clusters in SUSTAIN. With experience, participants learn to associate particular category labels with striatal units in different regions of stimulus space. A major difference, though, is that the location of striatal units is fixed, whereas clusters in SUSTAIN are created according to experience in a more flexible fashion.

1.7.5 Rational Model

Anderson’s rational model (Anderson, 1991) is a Bayesian clustering model that attempts to capture statistical regularities in the environment. In this model, internal subcategories or partitions are created by probabilistically grouping objects that share certain features, similar to a prototype. Several partitions can then represent a single category in the world. When encountering a new object, the model calculates the probability of classifying that object as a member of a category. The similarity of the object to the central tendency of the category is then weighted by how likely the category label is associated with other objects in that partition.

The rational model is conceptually similar to SUSTAIN in that it begins with a single cluster and adds additional clusters as needed. The difference lies in the goal of the model. SUSTAIN is concerned with recruiting clusters based on prediction failures or surprise events whereas the rational model attempts to determine the statistical structure of the world independent of the task at hand.

1.7.6 Comparing computational models of categorization

While SUSTAIN and Anderson's rational model have had a great deal of success at accounting for a wide variety of behavioral data, these are purely cognitive models that make no claims about neurobiological reality. This stands in contrast to the COVIS-based SPC model, which is motivated by the neurobiological properties of the basal ganglia. A major advantage to using neurobiology to constrain cognitive models is that it prevents one from building computational components into the model not supported by the neurobiology of the system. In addition, a consideration of the neurobiological plausibility can help differentiate between models that are able to account for the same pattern of results, but make different processing assumptions.

In Chapters 4-5, a new multi-system model of category learning that attempts to instantiate the cognitive operations active during learning is assessed with behavioral and imaging data of a category learning task thought to recruit the action of both RB and II systems. The advantage of this new model over existing models of RB and II categorization are that it attempts to account for the interaction between these learning systems and the modules of the model are intended to correspond to particular neural structures in the brain.

CHAPTER 2: NEURAL CORRELATES OF CATEGORY LEARNING

2.1 Overview

While neuropsychological and experimental behavioral work supports the notion that multiple memory systems are engaged in different types of category learning, fMRI provides a window into the functioning of a normal healthy brain during the learning process. Two fMRI studies of RB and II category learning that directly address the predictions made from the multiple memory systems theory are discussed below. In short, we observed a dissociation between the MTL and caudate in both studies when subjects were successfully utilizing an RB or II strategy, respectively.

2.1 fMRI Experiment 1: Testing COVIS with fMRI

2.1.1 Background

Previous neuroimaging studies have demonstrated the involvement of multiple neural systems in category learning, but no study had attempted to directly compare the neural correlates of RB and II learning strategies. Specifically, a clean dissociation between the RB and II systems had not been demonstrated. In a recent study (Nomura et al., 2007), we implemented the RB/II categorization task (Maddox et al., 2003) in a 3.0 Tesla scanner. In this study, participants learned two categories of sine wave gratings with the categories structured to encourage either a RB or II strategy. Gratings varied in spatial frequency (thickness of lines) and orientation (tilt of lines) according to the underlying category structure. In both groups, participants had to differentiate between two categories that were defined by a boundary line. This boundary was such that category membership in the RB group was defined only by the frequency of the lines (Figure 1.2A), but in the II group, frequency and orientation information needed to be combined in order to determine category membership (Figure 1.2B). In these figures, each point represents a distinct sine-wave stimulus and the colors indicate the different categories. Examples of category members are also shown. The consequence of these differing category structures is that successful RB categorization is done most effectively with a verbalizable, explicit strategy and II categorization with a non-verbalizable, implicit strategy.

2.1.2 Methods

2.1.2.1 Participants

Thirty-four healthy, native English-speaking, right-handed adults (15 males, 19 females) of mean age 23 (range 18 – 30) were recruited from the Northwestern University community for participation in this study. All participants gave informed consent according to procedures approved by the Northwestern University Institutional Review Board and were compensated for their time. Participants were randomly assigned to either the RB (N=17) or II (N=17) group. One RB participant was eliminated due to poor quality EPI data and 8 (4 RB, 4 II) were eliminated due to an inability to learn the categories (failing to meet a performance criterion of 60% correct on the final block).

2.1.2.2 Materials

Stimuli were circular sine wave gratings (Gabor patches; see Figure 1.2) that varied in spatial frequency (thickness of lines) and orientation (tilt of lines) as in Maddox et al. (Maddox et al., 2003). Participants were instructed to place each stimulus into one of two categories and to try to learn these categories over time based on the feedback given after each trial. The only difference between the RB and II groups was in the boundary that defined the categories. The stimulus space for both the RB and II groups can be thought of in 2 dimensions, spatial frequency on the x-axis and orientation on the y-axis. For the RB group, the stimuli were divided into categories based on a vertical decision boundary such that category membership depended only on the spatial frequency of the sine wave grating (Figure 1.2A). For the II group, the categories were defined by a diagonal decision boundary that required integration of spatial frequency and orientation information (Figure 1.2B).

2.1.2.3 Procedure

On each trial, a fixation cross was presented for 750ms followed by a single stimulus that was presented for 2 sec and during this time, participants indicated to which category they judged the stimulus belonged. Stimulus offset was followed by a 500 ms visual mask and feedback for the participant's choice ("Right", "Wrong") was shown for 750 ms. Participants were warned ("Time") if they had not made a response during the 2 sec the stimulus was on the screen (see Figure 1.2C). A total of 320 categorization trials were performed by each participant divided amongst 4 80-trial blocks. An equal number of fixation-only trials were pseudo-randomly interspersed between stimulus trials to maximize the separability of the measured hemodynamic response.

2.1.2.4 MRI acquisition

fMRI data were collected using a Siemens TRIO 3.0 T MRI scanner equipped with a transit/receive head coil while participants performed the categorization task. Whole-brain, gradient-recalled EPI (35 axial 3 mm slices, 0 gap) were collected every 2 sec (TR=2000; TE= 25 ms; flip angle = 78°; 22 cm FOV; 64x64 acquisition matrix; resulting voxel size = 3.44 x 3.44

x 3 mm) for 330 volumes in each of 4 scans. For anatomical localization, high-resolution, 3D MP-RAGE T1-weighted scans (voxel size = 0.859 mm x 0.859 x 1 mm; 160 axial slices) were collected for each participant following the functional runs.

2.1.2.5 Data Analysis

The functional images were first co-registered through time to correct for motion using a 3D alignment algorithm (Cox, 1996). Voxels with low signal (Cox, 1996) or excessive sudden signal change were eliminated (>30% in 2 sec) and the EPI data were smoothed (6.9 mm FWHM Gaussian kernel). Data were transformed to standard stereotactic space (MNI 305; (Collins et al., 1994)). Estimates of trial-locked evoked activity were made for the period of 4-12 seconds after stimulus onset to account for hemodynamic delay with overlapping responses deconvolved via a general linear model. Brain regions in which activity was associated with successful categorization were identified by comparing activity during correct and incorrect trials for each participant. Within each group, brain regions that exhibited consistently greater activity during successful categorization was identified by a second-pass random effects analysis. Monte Carlo simulation identified a reliability threshold of $t>4.5$ ($p<.01$ uncorrected) for all voxels in clusters of at least 300 mm^3. This method estimates the false positive rate for the study by creating random noise data that matches the mean and standard deviation of each voxel for each participant, maintaining the spatial structure of the data. The noise data is then subjected to exactly the same analysis as the real data (including spatial smoothing and the two-pass random effects model) and a statistical threshold is identified for which fewer than .05 false positive clusters are identified anywhere in the brain for the study.

In addition to the whole-brain analysis, the ability to identify anatomical boundaries for two critical regions hypothesized *a priori* to be important for category memory enabled a specific region of interest (ROI) analysis in the hippocampus and the caudate. For each participant, ROIs were drawn following anatomical boundaries that are visible on structural MRI. The MTL ROIs were drawn using boundaries that are described elsewhere (Insausti et al., 1998; Reber et al., 2002). The caudate ROIs were drawn according to known neuroanatomical boundaries separating the caudate from the surrounding white matter and ventricles. Each individual's ROIs were then aligned using the ROI alignment (Reber et al., 2002) method described in Stark and Okada (Stark and Okada, 2003). This method optimizes regional alignment at the expense of whole-brain alignment allowing for more precise localization and enhanced statistical power. Of particular interest was to test whether these two regions play different roles in RB and II categorization, i.e., whether there was significantly different activity associated with successful categorization in the RB and II groups. Separate reliability thresholds for contrasts between the participant groups within the ROIs were identified by additional Monte Carlo simulations (the MTL ROI volume was 21,500 mm^3, the caudate ROI was 11,000 mm^3; note that this method matches the shape as well as providing a "small volume" correction for the ROI volumes). Within the targeted ROIs, an alpha level of .05 is met by requiring clusters for which each voxel exhibited $t(24)>2.0$ to be at least 700 mm^3 in volume for the MTL, 600 mm^3 for the caudate.

2.1.3 Results

2.1.3.1 Behavioral performance

For both groups of participants who learned the task, performance was above chance in all runs, and the groups demonstrated similar learning curves (Figure 2.1). Learning across runs was reflected in a significant linear trend (F (1,23) = 111.6, $p < 0.05$). Mean accuracy averaged across all 4 runs for the RB group was 70.8% (SE = 1.83) and for the II group was 74.9% (SE = 3.03). RB and II accuracy was not significantly different across all 4 runs (F (1,23) = 1.25, n.s.), thus one cannot attribute differences observed in functional activity to task difficulty. Reaction-time also did not differ significantly between the groups (F (1, 23) = 3.162, n.s.) across all 4 runs.

2.1.3.2 Neuroimaging

2.1.3.2.1 Whole brain Analysis

In Figure 2.2, activity evoked by correct categorization was compared to that evoked by incorrect categorization to identify brain areas associated with successful categorization. The normalized atlas coordinates of the center of each of these clusters for the RB and II groups are listed in Table 1. In the RB group (Figure 2.2, top) successful categorization was associated with activity bilaterally in the MTL, body of the right caudate, anterior cingulate and medial frontal gyrus. In the II group (Figure 2.2, bottom), activity was observed bilaterally in the body and tail of the caudate for successful categorization.

There was a great deal of common activity between the groups that performed RB and II categorization. No differences were observed in total trial-locked activity (for all trials) between the groups, nor were there any global differences in total activity for correct categorization trials alone at a corrected threshold across the whole brain.

2.1.3.2.2 Region-of-interest analysis

The involvement of the MTL and caudate in RB and II categorization was examined with targeted anatomical ROI analysis to improve sensitivity to group differences in these crucial regions in which we had prior hypotheses about differential activity. Within the MTL ROI, the effect of successful categorization was significantly larger for the RB than the II group in the left anterior hippocampus. Correct RB responses evoked relatively greater activity than incorrect RB responses while the II responses did not vary with success. It should be noted that the general profile of activity in the anterior hippocampus was a negative deflection that occurred for all stimuli (see Figure 2.3C), whether the trial reflected correct or incorrect categorization. The RB success effect was an increase that effectively reduced the size of this deactivation. This type of trial-locked deactivation has been reported previously and interpreted as a familiarity signal during recognition memory (Henson et al., 2003) or as an indication that MTL activity may be elevated during "easy" or rest baseline conditions (Law et al., 2005), e.g., during the fixation periods interspersed among trials here. While it is tempting to suggest that assessing the familiarity of stimuli is involved in RB categorization, the size of this deactivation was not reliably different for RB and II categorization (although success did not affect MTL activity for II categorization) and further speculation should await a better understanding of the conditions under which trial-locked deactivations occur during memory tasks in the MTL. Regardless of the reason for the deactivation, the MTL activity was correlated with RB success, implying a role for the MTL memory system during the RB task.

In contrast, successful II categorization was associated with greater differential activity in the right body of the caudate compared with successful RB categorization (Figure 2.3B). In this region, increased activity was identified on all trials (Figure 2.3D), particularly correct trials, but the differential activity to successful categorization was reliably larger for II participants, suggesting that caudate activity is correlated with successful II categorization. Both targeted ROIs exhibit some commonality in the overall pattern of activity, possibly reflecting the competition between two simultaneously active categorization systems. However, the difference between activity on correct and incorrect trials indicates that the MTL activity is associated with successful RB categorization while the body of the caudate is associated with successful II categorization.

2.1.4 Discussion

Neuropsychological data has not provided a clear dissociation between the two types of category learning, likely due to interactions between these systems either through overlapping brain areas (e.g., the head of the caudate) or competition between the systems. However, functional neuroimaging provides strong support for the dissociation between the RB and II category learning systems. Several studies using fMRI to examine category learning have supported the idea that the basal ganglia, particularly the caudate, play a crucial role in category learning (Seger and Cincotta, 2005, 2006). Comparison of success-correlated evoked activity during category learning here (Nomura et al., 2007) shows that the roles of the MTL and the caudate can be doubly dissociated for RB and II category learning. This result was obtained using an alignment technique (ROI-AL) in the MTL and caudate that was previously only utilized in the MTL (Stark and Okada, 2003), which perhaps explains the lack of such an activity pattern in these areas previously. These findings all support the principle claims of COVIS that these two types of category learning are supported by separate neural systems with different operating characteristics.

It is notable that the peri-stimulus time course (Figure 2.3C) of evoked activity in the MTL cluster reflects a downward deflection for all trials. Unlike previous studies in which MTL deactivation was interpreted as reflecting competition with the basal ganglia, we observed deactivation in the MTL in conjunction with a success effect (relatively more activity for successful RB judgments than unsuccessful ones). An alternate interpretation for the deactivation in the MTL is that it is a familiarity signal as has been observed in a number of studies of old/new recognition memory (Henson et al., 2003). In our categorization tasks, the sine-wave stimuli are strikingly similar to each other both between and within categories, so the stimuli may seem familiar after just a few trials.

Observed deactivations often raise questions about the relationship of metabolism (measured by fMRI) and neural activity. However, in a recent study of epileptic patients prior to surgical resection, researchers recorded from neurons in the hippocampus and surrounding cortex while participants were responding to either novel or familiar stimuli (Viskontas et al., 2006) and found that familiar stimuli induced decreased firing rates such that activity was deflected below baseline. This observation provides strong evidence that the downward deflection seen in the hippocampus in fMRI studies can arise from an inhibitory mechanism for familiar items at the neural circuit level.

Although both types of category learning appear to evoke this familiarity signal, successful RB categorization seems to induce an additional relative increase in activity, creating a greater separation between correct and incorrect trial evoked activity. Overall, the pattern of activity in the MTL suggests the operation of two opposing processes. There is a tonic deactivation to each trial reflecting familiarity; and an increase in activity for successful trials, possibly reflecting successful recollection. The reduced deactivation for the II condition suggests one of two things: either the familiarity effect is attenuated for II or that the process associated with increased the activity for successful RB is active on all II trials regardless of the feedback.

For the II group there was greater differential activation in the right posterior caudate body than in the RB group (Figure 2.3B). The peri-stimulus time course (Figure 2.3D) shows that activity in the caudate increases on each trial, but the difference between correct than incorrect trials is greater in the II group. Despite the differential activity seen here between the MTL and caudate ROIs for the RB and II conditions, the similarity in the overall pattern of peri-stimulus time course activity observed suggests that both types of learning may be active simultaneously in both conditions, but that the one better suited to learning the category is more involved in successful performance. This may in part explain the lack of hypothesized differential activity in frontal cortex. It should also be noted that this subtle difference is one that is unlikely to be detected by a typical whole- brain voxel-based analysis, hence the need for the ROI analysis.

The results here indicate that successful RB category learning is associated with increased activity in the MTL, reflecting explicit knowledge of the category. In contrast, successful II categorization is associated with increased activity in the posterior caudate, reflecting the critical role of this area in feedback-oriented implicit category learning. While this study revealed a dissociation between components of two different types of memory, it was unable to reveal additional components of the hypothesized RB and II category learning networks proposed in the COVIS model.

2.2 fMRI Experiment 2: Increased within-category variance

2.2.1 Background

The dissociation between the MTL and caudate observed in Experiment 1 provides strong support for the multiple systems view of categorization. The aim of the second fMRI experiment was to address an alternative explanation for this dissociation. Namely, in experiment 1, the inter-category variability was purposefully mismatched between the RB and II groups, to ensure similar group performance. That is, the RB stimuli were closer to the category boundary line (less discriminable) than the II stimuli, making the tasks equally difficult. It could be argued then, that the activation differences observed for RB and II category learning were due to the degree of discriminability of the stimuli rather than differences in learning strategy. Specifically, since the RB stimulus set had more similar category members, it may have required a stronger, more specific representation than the highly discriminable II stimulus set, and this representation difference was responsible for the dissociation in activity between the MTL and the caudate. The purpose of Experiment 2 was to address this alternate hypothesis by matching the within-category variance between groups. The increased variance within the RB and II category structures also enabled analyses of trial activity close and far from the decision boundary to assess the effects of distance both on reaction time and functional activity.

Experiment 2 followed the same basic paradigm as Experiment 1 with an additional manipulation of the stimuli in both the RB and II conditions so they receive both “low-discriminable” (close) and “high-discriminable” (far) stimuli. Close stimuli were added to the previous II stimulus set and far stimuli to the previous RB stimulus set (Figure 2.4). This manipulation enabled us to replicate the findings of Experiment 1 by analyzing the activity associated with the close RB stimuli and the far II stimuli (we expected to see MTL and caudate activity respectively). Further, we can extend these findings by examining the activity associated with the new far RB and close II stimuli. Specific contrasts compare the close and far stimuli within the groups, so any difference in activation pattern can be attributed to the difference in discriminability.

If discriminability serves to distinguish between the systems, in both groups, we would expect that the close stimuli would be associated with MTL activity and the far stimuli with caudate activity. However, if the degree of discriminability has no affect on the categorization process (as we predict), all stimuli should result in activation of the MTL in the RB group and the caudate in the II group. In the behavioral data, because the far stimuli in the RB structure represent more salient examples of category membership than those close to the boundary, we expect RB performance to exceed II performance. In contrast, the close stimuli provide no more information to the II group, thus performance should be the same as in Experiment 1.

Even with this predicted change in performance, we expect similar activation consistent with our hypothesis that the MTL supports RB categorization and the caudate supports II categorization. Since Experiment 1 and 2 share some common features, we can directly compare these overlapping conditions to look for important changes. Specifically, by comparing the activation to close stimuli in the RB group to the previous RB group results and the activation to far stimuli in the II group to the previous II group results, we will be able to determine whether the pattern of activation changed with this new design, reflecting a possible strategy change. Since the hypothesis is that the learning strategy is based on the category structure rather than the discriminability of the stimuli, we expect to replicate the results of Experiment 1. That is to say, the presence of far stimuli in the RB group and close stimuli in the II group should not affect the process by which the categories are learned.

2.2.2 Methods

2.2.2.1 Participants

Thirty-three healthy, native English-speaking, right-handed adults (10 males, 23 females) were recruited from the Northwestern University community for participation in this study. All participants gave informed consent according to procedures approved by the Northwestern University Institutional Review Board and were compensated for their time. Participants were randomly assigned to either the RB (N=11) or II (N=22) group. Two II participants were eliminated due to poor quality EPI data (due to subject movement). We recruited twice as many II subjects because initial modeling analyses revealed a sub-set of II subjects using an RB strategy (discussed in Chapter 3).

2.2.2.2 Materials

The task here was the same as in Experiment 1. Again, there was an RB and an II group that observed sine wave gratings that varied in frequency and orientation. In both groups, the stimuli came from 2 different categories, where the category boundary was the only thing that differed between the groups. The RB stimuli were defined by a vertical boundary where category membership depended only on frequency, and the II stimuli were defined by a diagonal boundary that required an integration of frequency and orientation information. There was an additional modification of both stimulus sets in that there were high-discriminable (far) stimuli added to the existing low-discriminable RB set and low-discriminable (close) stimuli added to the existing high-discriminable II set (Figure 2.4). Far stimuli were defined as being far from the category-bound, and were thus easily differentiated from the other category. Close stimuli were more proximal to the category-bound, making the category membership more difficult to discern. Mathematically, the division between 'far and 'close' was determined based on the mean frequency and orientation values of the category. These values were determined for both categories of each individual's category structure, and the trials were marked accordingly.

2.2.2.3 Procedure

As in Experiment 1, participants were placed in a 3.0 Tesla GE MRI scanner and observed images on a rear-projection screen via a mirror mounted above their eyes. Before beginning, all participants regardless of group were instructed that they should categorize the stimuli while still on the screen by pressing one of the buttons they held in their hands. In this event-related design, fixation trials were pseudo-randomly interspersed between the trials to maximize the separability of the hemodynamic response to stimulus trials. On fixation trials, participants were instructed to fixate on the crosses without making a response. On the stimulus trials, they made a response by pressing either the left or right button indicating one or another category, and then received feedback. The procedure was the same for both the RB and II groups, the only difference was in the category boundary between groups. All participants were instructed to make a category judgment while the stimulus was on the screen by pressing a hand-held button in either their right or left hand. There were 4 functional runs, each run consisting of 80 trials and 80 blank fixations. The anatomical localization scan followed the functional runs.

2.2.2.4 MRI acquisition

fMRI data were collected using a GE 3.0 T MRI scanner equipped with a transit/receive head coil while participants performed the categorization task. Whole-brain, gradient-recalled EPI (40 axial 3 mm slices, 0 gap) were collected every 2 sec (TR=2000; TE= 25 ms; flip angle = 78°; 22 cm FOV; 64x64 acquisition matrix; resulting voxel size = 3.44 x 3.44 x 3 mm) for 326 volumes in each of 4 scans. For anatomical localization, high-resolution, 3D MP-RAGE T1-weighted scans (voxel size = 0.859 mm x 0.859 x 1 mm; 160 axial slices) were collected for each participant following the functional runs.

2.2.2.5 Data analysis

Preprocessing and statistical analysis of the data were performed with a collection of software based on AFNI (Cox, 1996) as in Experiment 1. The functional images were co-registered through time to correct for motion, normalized to MNI stereotactic space, and spatially smoothed. For each participant, voxels were fit to a general linear model function that is based on blocking stimuli to track activity that changes as a result of processing specific types of trials. We then created separate vectors marking the different trial types—correct, incorrect, close, far. In addition to looking at the differential activation to correct as compared to incorrect trials in both the RB and II groups, we also looked at the difference in activation to close and far trials both within and between groups. This allowed us to draw conclusions about the effect of discriminability of the stimuli on the neural correlates of RB and II category learning.

Based on a priori hypotheses about specific neuroanatomical areas, we also used a region of interest (ROI) analysis in the hippocampus/parahippocampal gyrus and the caudate nucleus of the basal ganglia. These ROIs were hand-drawn in each subject using anatomical boundaries visible on a structural MRI scan (as in Experiment 1). The boundaries used to draw the hippocampal ROIs are discussed elsewhere (Insausti et al., 1998; Reber et al., 2002). The boundaries used to draw the caudate ROIs were based on known neuroanatomical boundaries of the surrounding ventricles and white matter. These ROIs were drawn independent of the location of functional activity.

2.2.3 Results

2.2.3.1 Behavioral performance

For both groups of participants, performance was above chance in all runs, and the groups demonstrated similar learning curves (Figure 2.5). Learning across runs was reflected in a significant linear trend (F (1, 29) = 24.716, $p < 0.05$). Mean accuracy averaged across all 4 runs for the RB group was 77.3% (SE = 0.034) and for the II group was 66.6% (SE = 0.055). The RB group accuracy was significantly greater than the II accuracy across all 4 runs (F (1,29) = 12.5, $p< 0.05$).

2.2.3.1.1 Distance analysis

The accuracy of the subject groups differed as a function of distance of the stimuli from the decision bound (Figure 2.6). The mean accuracy for close trials was 63.7% (SE=0.011) and for far trials was 77.5% (SE=0.02). This difference was significant (F=67.948, $p<0.01$). On close trials, RB subjects were 66.6% (SE=0.018) accurate and II subjects were 60.8% (SE=0.013) accurate. On far trials, RB subjects were 80.3% (SE=0.032) accurate and II subjects were 74.7% (SE=0.024) accurate. This distance effect was significant within both the RB and II subject groups, so the interaction was not significant (F=0.005, n.s.).

2.2.3.1.2 Reaction Time analysis

There were reaction time (RT) differences between the groups as well as within the different trial types (Figure 2.7A). The mean RT for RB subjects was 1.115 sec (SE=0.028) and for II subjects was 1.159 sec (SE=.021). Overall, RB subjects were significantly faster than II subjects (F (1,29)=4.523, p<.05). When the trials are broken down based on whether they were correct or incorrect, for the RB subjects, the mean RT for correct trials was 1.039 sec (SE=.031) and incorrect trials was 1.191 sec (SE=.027). For II subjects, the mean correct RT was 1.126 sec (SE=.023) and incorrect RT was 1.191 sec (SE=.02). For correct trials, the RB group was significantly faster than the II group (F (1,29)=4.945, p<0.05), but on incorrect trials, the groups did not differ (F (1,29)=0, n.s.). There was a significant interaction between the groups (F (1,29)=14.094, p<0.05).

When trials were separated based on distance to the boundary, there were RT differences between ‘close’ and ‘far’ trials (Figure 2.7B) within and between groups. In the RB group, far trials were faster than close trials (F (1,20)=3.4, p<0.05). The II group did not have any RT differences between close and far trials (F (1,38)=0.596, n.s.). There was a significant within-subject interaction between distance and group (F (1,29)=14.396, p<0.05). There was also a significant between-subject interaction across group (F (1,29)=5.26, p<0.05).

2.2.3.2 Neuroimaging

2.2.3.2.1 Whole-brain analysis

In Figure 2.8, activity evoked by correct categorization was compared to that evoked by incorrect categorization across all 4 runs to identify brain areas associated with successful categorization. In the RB group, the significant regions of activity were all greater for incorrect than correct trials (Figure 2.8A). Specifically, the insula was active bilaterally and regions in the right frontal cortex were also more active during incorrect than correct trials. In the II group (Figure 2.8B), activity in the body and tail of the caudate was greater for correct than incorrect trials. Also, the right PFC was more active for incorrect than correct trials in II subjects. When the successful categorization activity was compared directly between the groups, there were no observed significant differences across all 4 runs.

The designation of close and far stimuli allowed for additional contrasts of activity. Within the RB group, the contrast of correct and incorrect close trials revealed a number of significant regions of activity (Figure 2.9 and Table 2.2). The majority of activation in this group was greater for incorrect than correct trials with the exception of the cuneus and right parietal cortex. Bilateral activation in the insula and the frontal cortex was observed greater for incorrect close trials. There were also medial frontal and some posterior parietal lobule activations. During the far trials, there was less activation overall, but those areas that were significantly different were in the insula, frontal and parietal cortices. Numerically, there were fewer incorrect far trials in the RB group, so it may not be fair to compare to the close trial activation maps. Additional clusters of activity are listed in Table 2.2.

Within the II group, significant differences between correct and incorrect trials were only observed for close but not far trials (Figure 2.10). As in the RB group, the insula and frontal cortices were more active for incorrect than correct close trials. The body of the caudate and right parietal cortex were more active for correct than incorrect close trials in the II subject group. Additional clusters of activity are listed in Table 2.3.

2.2.3.2.2 Region-of-interest analysis

The involvement of the MTL and caudate in RB and II categorization was examined using a targeted anatomical ROI analysis of these structures to improve sensitivity to group differences in these crucial regions in which we had prior hypotheses about differential activity. We observed a slightly different activity pattern than that seen in Experiment 1 where RB successful categorization activity was greater in the left anterior MTL and II was greater in the posterior body of the caudate. Here, both the RB and II subject groups demonstrated successful categorization differences within the MTL, in slightly different regions. In the RB subject group, successful categorization activity was greatest in the left anterior MTL (Figure 2.11A), but was also present in the right anterior MTL. The II group showed a similar successful categorization difference within the left anterior MTL, but the cluster was smaller than that in the RB group (Figure 2.11C) and extended more laterally. There were no significant regions of activity within the MTL ROI for the double subtraction between the RB and II subject groups, presumably due to the presence of similar activity in the single group subtractions. In the caudate ROI, however, there were significant differences between the RB and II subject successful categorization activity. In Figure 2.12 (bottom), three significant clusters of activity were detected in this double subtraction. The top portion of Figure 2.12 shows the single group caudate ROI differences for successful categorization to demonstrate the difference in direction of the activation between groups. The RB group was more active in these regions during incorrect than correct trials, whereas the II group was more active for correct than incorrect trials. The double subtraction reveals the greatest difference in categorization success in the bilateral body of the caudate (Figure 2.12 bottom).

We also examined activity associated with perceptual discriminability within the ROIs to determine whether the close and far trials elicited differential engagement of the MTL and caudate. As predicted, there was no difference between close and far successful categorization-associated activity within the MTL and caudate. For the II subjects, a cluster reflecting the difference between correct close and incorrect close trials was observed in anterior MTL (Figure 2.13 bottom left) and this same area was also revealed in the contrast between far correct and far incorrect trials (Figure 2.13 bottom right). For the RB subjects, the successful categorization difference was evident in close trials, but not in far trials (Figure 2.13 top). However, the relative lack of incorrect far trials may have precluded an appropriate subtraction here. In the caudate ROI, the distance to the bound had no effect on activity in posterior caudate (Figure 2.14) in either the RB or II group.

2.2.4 Discussion

The data from the second fMRI experiment both replicated the findings of Experiment 1 and provided some evidence that the dissociation between the RB and II systems was related to strategic differences rather than to the specificity of the category representation. The reaction-time (RT) differences between the RB and II groups suggest that the differences in the construction of the categories placed different cognitive demands on the subjects. The task appeared much easier for the RB than the II group here, as their RTs were faster and accuracy was higher overall than in the II group. When the trial types were separated into ‘close’ and ‘far’, for the RB group, those stimuli located close to the category boundary were more difficult and elicited slower RTs than those that were far from the bound. For the II group, however, there was no RT difference between trial types, suggesting that the distance from the category boundary was not correlated with ease of categorization for the II stimuli.

Due to this change in stimulus distributions, as predicted, the RB group learned more quickly and reached a higher level of accuracy than the II group. The presence of category members located far from the category bound provided the subjects with a clear indication of the relevant stimulus feature (frequency). The presence of far stimuli did not equivalently affect performance in the II group, as their performance was overall lower than the RB group. This performance difference was predicted, and a necessary result of matching the within-category variance. In the previous study (Nomura et al., 2007), performance was matched by reducing the distance of the RB stimuli from the boundary.

The fMRI data revealed a similar pattern of activity in the focused region-of-interest analysis to that in Experiment 1. That is, successful RB categorization was associated with greater activity in the MTL than II categorization and II categorization elicited greater activity in the posterior caudate than RB categorization (Figure 2.12). An additional observation in Experiment 2 was the presence of successful RB categorization activity in the MTL for II subjects. While the utilization of an RB strategy is not optimal for acquiring an II category structure, the difficulty of the task may have encouraged the adoption of a non-optimal strategy in some II subjects. The model-fitting techniques discussed in later sections (Chapters 3-6) have demonstrated that this is the case.

Within these MTL and caudate clusters of activity, there were several important differences between experiments in the peri-stimulus timecourses. In the MTL (Figure 2.11B), the RB group activity, as a whole was less negative than the II group activity yet the difference between correct and incorrect trial activity was larger in the RB than the II group. In Experiment

1, the opposite effect was observed where RB trials were more negative than the II trials (Figure 2.3 C). This finding is puzzling, yet may be explained by the involvement of the RB system in II subjects. The difficulty of the II learning task in Experiment 2 may have led a number of the II subjects to adopt an RB strategy and thus recruited the operation of the MTL. Mathematical models fit to these II subject runs (Chapter 3) supports this interpretation.

In the caudate, the double subtraction between successful categorization in RB and II subject groups reveals a number of significant clusters of activity. The activity contributing to these clusters differs from that observed in Experiment 1. Here, the incorrect trials serve to drive the group differences. In all clusters, incorrect trials in the RB group elicit greater activity than incorrect trials in the II group. The correct trial activity in both groups does not differ. The apparent increase in activity for incorrect trials in the RB group and reduced activity for incorrect trials in the II group suggests that negative feedback is being processed differently in these subjects.

In the whole brain activity, there was less consistency between the studies, but there are a number of reasons that could account for this. As mentioned previously, the category structures differed in an important and deliberate manner between the two experiments. Here we matched the mean distance of the stimuli from the bound between the RB and II groups whereas in Experiment 1, the RB categories were less discriminable than II. Consequently, the groups differed in accuracy. This discriminability difference may also have changed the cognitive strategy employed by participants, particularly in the II group.

Consistent with our hypothesis, we observed similar patterns of activity for stimuli ‘close’ and ‘far’ from the category boundary within the anatomical ROIs. For the RB group those stimuli located closer to the category bound elicited similar MTL activity to those trials that were farther from the bound, regardless of whether the response was correct or incorrect. Similarly, in the II group, caudate activity was observed for both close and far trials. In addition to replicating the effects observed in Experiment 1, here we have demonstrated that discriminability of the stimuli cannot account for the differential activity observed during RB and II category learning. Rather, the different learning processes required for RB and II categorization preferentially recruits the MTL and caudate-based memory systems.

CHAPTER 3: MATHEMATICAL MODELING OF RB AND II STRATEGY USE

3.1 Overview

Mathematical modeling can provide an alternative method for identifying strategy-use rather than the more typical accuracy-based analyses such as those utilized in Chapter 2. Particularly in tasks such as those used in the previously described fMRI studies, category structures that can be described with a mathematical formula (e.g. linear decision bound of a particular frequency value) are well suited to this approach.

In the fMRI data discussed in Chapter 2, all analyses compared correct to incorrect trial activity to isolate areas associated with successful category learning. While this was effective for the group analysis, the mathematical models here successfully identified blocks of trials in individual subject data that were fit best by either a RB or II strategy independent of the stimulus distribution assigned to those subjects. This re-grouping allowed for the isolation of activity associated specifically with RB or II category learning. For example, those blocks of trials better fit with an RB than an II model were associated with activity in the PFC and ACC and likewise, best fit II blocks were associated with extrastriate visual cortex activity.

Mathematical models that objectively characterize the behavior of the subject are able to identify strategy use independent of accuracy assessments. This approach proved particularly useful here as we did not explicitly instruct participants toward using one strategy over another. Indeed, modeling of the data from both Experiment 1 and 2 revealed that a number of II subjects were using an RB strategy early in learning and transitioned to II learning as they successfully acquired the category representation. Using the model-based analyses to isolate participants employing different strategies provides a more rigorous window in to potential brain activation and performance differences. The following chapter discusses 2 mathematical models (RB and II) that were used to fit the behavioral data collected in the previously discussed fMRI studies. Functional activity during the model-identified 'best fitting' runs is examined and discussed in the context of the COVIS theory.

3.2 Methods

3.2.1 RB and II model parameters

Here we have instantiated RB and II decision-bound models derived from general recognition theory (GRT; (Ashby and Townsend, 1986)), which is a multivariate generalization of signal-detection theory. On each trial, it is assumed that the percept can be represented as a point in multidimensional psychological space. Decision bound theory then assumes that each participant partitions the perceptual space into response regions by constructing a decision boundary. On each trial, the participant determines which region the percept is in, and produces the associated response. Despite this deterministic decision rule, decision bound models predict probabilistic responding because of trial-by-trial perceptual and criterial noise (Ashby, 1992).

Following previous work (Maddox et al., 2004b), both the RB and II models were fit separately to the behavioral data (for Exp 1 and 2) for each of the four 80-trial blocks for each participant. In both the RB and II models, 2 parameters were used to fit the data. The first was an intercept that reflected the location of the decision boundary in stimulus space. The second was a perceptual shaping parameter that reflects the distinctiveness of the categories within the stimulus space. When this parameter value is high, it represents the difficulty in discriminating two perceptually similar stimuli, but when the value is low, the category representation is more precise. With learning, the magnitude of this parameter should decrease, reflecting an increased ‘sharpness’ of the representation of the categories, a defining feature of perceptual learning.

The RB model assumes a vertical decision boundary (in stimulus space) reflecting the use of a rule dependent on a single stimulus dimension (e.g. frequency). The II model assumes a decision boundary with slope equal to 1.0 (i.e. a diagonal line reflecting integration of both dimensions). Thus, both the RB and II models have two free parameters (x-intercept and perceptual shaping parameter for RB and y-intercept and perceptual shaping parameter for II), which allows for a direct comparison of fit value between models. Each block of 80 trials was fit separately with both the RB and II model to identify the decision bound and perceptual shaping parameter that best accounted for the observed data.

3.1.2 Model fitting

Best fitting parameter values were identified by a downhill simplex method (Press et al., 1992). These RB and II models were fit to each run of each subject in order to identify the subject runs that were ‘best fit’ by a particular strategy. After sorting all model fits from high to low, only the top third from both RB and II model fits were included in the resulting fMRI analyses. In some cases, subjects contributed more than one run to the analysis. This classification of 80-trial blocks was used to organize the fMRI data collected previously (Experiment 1 and 2) and examine whether the model-identified periods of best strategy expression clearly identified additional elements of the neural networks for RB and II category learning. The increased within-category variance in Experiment 2 was intended to enhance the efficacy of modeling this data.

3.3 Results

3.3.1 Experiment 1 data

3.3.1.1 Mathematical models fit to behavioral data

Out of 25 subjects (4 blocks each), there were 95 useable blocks of data to model (lost 5 blocks due to noisy fMRI data), 44 in the RB group and 51 in the II group. A color-coded schematic of the best fitting model per run is shown in Figure 3.1. The blocks in red, magenta and pink represent the top, middle and low RB model fits, respectively. The blocks in navy, blue and light blue represent the top, middle and low II model fits, respectively.

Within the RB group, 38 out of 44 runs were best fit with the RB model. Within the II group, 39 out of 51 runs were best fit with the II model. There were a number of subjects that used a non-optimal strategy, particularly in the II group (e.g. II subj #12).

3.3.1.2 Mathematical model-driven reorganization of fMRI data

Figure 3.2 contrasts the top 15 subject runs that were best fit by the RB model with the top 15 runs best fit by the II model for correct trials only. The activity during correct trials was used because the well-fit runs tended to be those runs where there was a high degree of accuracy and few incorrect trials. The results of this analysis showed that in the right PFC, activity was greater for RB than II fit runs. In a right visual cortical region, greater activity was observed during well fit II than RB runs.

From the model-based fMRI analysis, we isolated two regions in the whole-brain activity that were not present in previous analyses. The model-identified mixture of RB and II strategy use within the RB and II groups may have masked the dissociation in these areas, highlighting the effectiveness of using these mathematical models to separate categorization behavior and corresponding functional activity. These results are consistent with COVIS's prediction the PFC is involved when one is correctly utilizing an RB strategy and the occipital cortex is involved when one is correctly utilizing an II strategy.

3.3.2 Experiment 2 data

3.3.2.1 Mathematical models fit to behavioral data

Out of 31 subjects, there were 120 useable blocks of data to model (lost 4 blocks due to due to noisy fMRI data), 43 in the RB group and 77 in the II group. A color-coded schematic of the best fitting model per run is shown in Figure 3.3. Thirty-seven out of the 43 RB subject runs were fit best with the RB model and 34 out of 77 II subject runs were best fit with the II model. Thirty-three out of 77 II subject runs were fit better with the RB than the II model. The remaining II subjects were fit better with the RB model on some runs and better with the II model on others, suggesting these subjects were particularly prone to strategy switching behavior.

3.3.2.2 Mathematical model-driven reorganization of fMRI data

The performance in Experiment 2 was lower overall than that in Experiment 1, so it was possible to contrast correct and incorrect trial activity. Figure 3.4A shows the result of this successful categorization subtraction within the top 26 subject runs that were best fit by the RB model. Areas associated with correct trial activity were the left anterior MTL, left superior frontal cortex, right parietal cortex and posterior cingulate. Similar MTL activity was also seen in the ROI analysis using the original experimental grouping of subjects (Figure 2.11A). That the same area was revealed in this modeling-based analysis of the same data suggests that the best fit RB runs here may have been driving the effect seen in the ROI.

There were also some areas that were more active during incorrect trials, particularly in PFC and medial frontal cortex. The pattern of activity related to correct and incorrect trials suggest that these trial types are being processed differently. One interpretation is that the MTL and frontal cortices are involved with maintaining the appropriate rule, whereas the inferior and medial PFC are involved when one receives disconfirming information. COVIS maintains that the RB system relies upon hypothesis testing, and it follows that incorrect feedback should be particularly important for updating the decision boundary.

This is one example where fMRI analysis can inform models of categorization. Given two models that make different predictions about feedback processing, the evidence shown here would support a model that assigns the processing of positive and negative feedback to different neural structures. Chapter 5 examines this hypothesis in more detail.

Figure 3.4B contrasts the activity associated with correct and incorrect trials within the top 14 subject runs that were best fit by the II model. During correct trials, there was activity in the left body of the caudate as well as in the left visual association cortex and parietal cortices. While the caudate involvement was expected, the visual cortical area was not present in the previous fMRI analysis. By only including subject runs that, based on their behavioral profile, were utilizing an II strategy, we've identified an additional component of the II categorization network. This area fits in with the COVIS theory of II learning where the visual association cortex acts to acquire a category representation in association with the posterior caudate. This region was also identified in the model-based analysis of Experiment 1, which suggests that the components of the II categorization network are similarly active despite the differences in stimulus sets between studies.

A specific example of the efficacy of the model in identifying the learning strategy independent of the imposed category structure is shown in Figure 3.5. This data comes from a single II subject run that was best fit with the RB model. This subject, despite receiving disconfirming feedback, persisted in using an RB strategy. The pattern of functional activity observed during this run is similar to that of the group of RB-modeled subjects (Figure 3.4A). Specifically, there is MTL and frontal activity for correct trials and inferior PFC activity for incorrect trials. The functional activity for a number of II subjects that showed this profile of RB strategy-use shows an overall pattern of activity that is much closer to the 'RB network' than the 'II network' (Figure 3.6A). Further, the presence of success-related MTL activity (Figure 3.6B) in these subject runs suggests that these particular II subjects may be contributing toward the activation seen in the MTL ROI in Figure 2.11C. Without a model-based trial sorting methodology, activity for these II subjects that are using an MTL-based RB strategy would be grouped in with that of the II subjects that are utilizing an II strategy.

3.4 Discussion

Functional neuroimaging provides an effective tool for testing predictions of neurally-inspired models of cognitive function such as COVIS. This tool can be made even more effective by incorporating DBT-based mathematical models that more effectively characterize strategy-use. As seen above, DBT provides a mechanism for examining the detailed behavioral data produced by a participant and conjecturing about the participant's mental state (e.g., whether the participant is depending on an RB or II strategy). Using this analysis to guide fMRI data analysis can test predictions of the COVIS theory. Comparing the participants who best exhibited an RB strategy with the best exhibited II strategy indicated additional areas of dissociation in frontal and occipital cortical areas, as predicted by COVIS. Note that because COVIS is a hypothesis about how DBT can be accomplished in the brain, there is no circularity here. If PFC areas had been associated with II category learning (or posterior areas with RB), the predictions of COVIS would have been contradicted. The successful utilization of the model-fitting technique shown here to isolate the most effective applications of RB and II strategies suggests that this method in combination with fMRI can be used to further identify the brain networks supporting these processes.

CHAPTER 4: PINNACLE: A NEW MODEL OF CATEGORY LEARNING

4.1 Overview

The observed dissociation between RB and II category learning in the MTL and the posterior caudate (Nomura et al., 2007) lends support to the notion that there are multiple systems in the brain capable of creating novel visual category representations. While we can speculate as to the specific role of each of these systems in category learning based on the known neurophysiology and operating characteristics of the MTL and caudate in other contexts, neuroimaging alone cannot address the question of what mechanism(s) drive learning in both these networks. Furthermore, the nature of interaction between the systems is unknown, and cannot be deduced from dissociations in fMRI data. The development of a model that instantiates a cognitive theory of RB and II category learning systems allows us to test more specific hypotheses about how the brain is able to acquire these relatively simple category structures. Using the model to make predictions in behavioral and fMRI data provides opportunities to test the theory in a specific, systematic manner.

The following chapter describes our theory of RB and II category learning, which stems from the COVIS theory, but is articulated through a cognitive model called PINNACLE (Parallel Interactive Neural Networks Active in Category LEarning). PINNACLE differs from the computational implementation of COVIS in a number of important ways, but the biggest difference is in the number of free parameters. Our approach to building this model was to start with the minimum number of parameters and to add additional ones when the model could not account for the observed behavior.

4.2 Background

The DBT model of categorization is a relatively simple method for describing boundaries in perceptual space that lead to visual category representations. It has typically been used in paradigms in which the experimental stimuli can be described in a small number of dimensions (usually 2). However, these models are limited in the sense that they say very little about the hypothesized cognitive operations necessary for RB and II category learning to co-exist within the brain. In contrast, sophisticated cognitive models contain strong hypotheses about representational structures but do not typically indicate how these structures are represented in neural systems. For example, it is not immediately clear how SUSTAIN or Anderson's rational categorization model proposes to implement cluster representations in the brain or how these change with experience.

The PINNACLE model aims to represent the cognitive state of the participant on every trial including the current best estimate of the RB system representation, the II system representation and the decision structure that adjudicates competition between the systems. The resulting model trial-by-trial predictions have the potential to enhance the analysis of functional neuroimaging data by identifying individual trials that are particularly clear examples of specific strategy use.

As the model becomes more detailed and specific, it also provides an opportunity to test its psychological reality by examining whether specific parameters have real neural correlates. An important goal in constructing PINNACLE was that every parameter and component be associated with a particular cognitive process. In those cases where a parameter is identified that cannot be mapped onto a cognitive or neural process, comparing the relative fit value of the data to that generated by a version without the parameter in question can evaluate its necessity. In this way we hope to avoid the danger of over-fitting the data with extraneous parameters.

4.3 Cognitive processes modeled in PINNACLE

The hypothesized organization of the neural circuits supporting RB and II category learning in Ashby's COVIS theory is outlined in Figure 1.3 and was discussed previously. While the PINNACLE theory is adapted from COVIS and thus implicates many of the same neural structures, the role of the MTL in supporting RB category learning and the posterior region of the caudate in supporting II category learning are emphasized in the PINNACLE model. Accordingly, the representation of the category structure in the RB system here is designed specifically to mimic the operation of the MTL. We hypothesize that RB categorization as a whole, though, depends on learning within a network consisting of the PFC, MTL, head of the caudate, and the ACC. In contrast, II categorization is thought to depend on learning in posterior cortico-striatal circuits between extrastriate visual cortex and the body/tail of the caudate. Both systems feed information to the PFC to guide motor responses about category membership.

4.3.1 RB system

The role of the RB system in the model is to represent the current decision bound of the participant. Since the category structures utilized here were based on linear decision boundaries, similar to the mathematical models discussed previously, the current model uses a vertical line to track the RB system's representation of the category structure. This line can be conceptualized as a verbalizable rule that shifts according to the current state of learning in the participant.

In terms of functionality, the RB system requires enough flexibility to enable switching amongst different verbalizable rules in accordance with feedback information. This requires the involvement of both attention and working memory processes in the PFC to not only hold the most recent rule in mind, but to compare the previously seen examples of a particular category to the current stimulus. Explicit memory processes should support this latter function in particular, facilitated by reciprocal connections between the MTL and PFC.

4.3.2 II system

Learning in the II system is highly sensitive to feedback information. As discussed previously, the nature and timing of feedback is essential to the operation of the II system. We hypothesize that learning in the II system is mediated by cortico-striatal loops through the posterior portions of the caudate of the basal ganglia to visual association cortex (Yeterian and Pandya, 1995). These loops have been considered by others (Houk et al., 2007) to be computationally powerful processing modules that allow for action selection. The projections from posterior visual cortices to PFC provide a pathway by which the category representation becomes associated with a particular verbal label. The dompaminergic input comes from neurons within the substantia nigra, and through a process similar to reinforcement learning, the most recently active synapses responsible for the correct response are strengthened (Wickens, 1990; Schultz, 1992).

4.4 Implementing PINNACLE theory into a cognitive computational model

The theory guiding the construction of PINNACLE assumes two learning systems that rely on MTL- and caudate-based memory compete on each trial to drive behavior. In this section, the mathematical operations that make up the RB and II systems of PINNACLE as well as the model fitting methodology are described. The category representations within the RB and II systems of PINNACLE can be thought of as analogous to the DBT-style mathematical models discussed in Chapter 3. What is novel about PINNACLE is that both systems are available to contribute to learning on every trial just as a healthy human is capable of using all available resources in a given learning task.

The process of evaluating the operation of PINNACLE involves several steps. The first is to demonstrate its ability to generate learning curves similar to those observed in the RB and II groups in Experiments 1 and 2. This step is important in establishing that the mechanics of the model not only enable learning, but that the progression of learning is similar to that of actual subjects. In the next step, we evaluate whether PINNACLE is capable of fitting individual subject behavior on a trial-by-trial basis. This means finding a set of parameters for each subject that produce the best account of that subject's response profile.

Through the model-fitting process, details about the model's structure regarding the number of free parameters of the model, the mechanism of interaction between systems and the appropriate feedback mechanism were addressed. In all these cases, model simulations applied to the available behavioral data provided a manner of testing competing hypotheses.

4.5 Model structure

The structure of the model is based on the hypotheses put forth by the COVIS theory of category learning. That is, similar to COVIS, the PINNACLE model contains 2 learning systems that compete on each trial to provide the output response. Based on feedback after each trial, the system that provided the most recent response updates its representation. A schematic of the model is shown in Figure 4.1. Where PINNACLE differs from COVIS is on the specific emphasis on the MTL and the caudate and the incorporation of neurobiologically plausible mechanisms for implementing the theory.

Our general approach in building this model was to constrain the number of parameters to those that we could relate to a specific psychological construct. One consequence of this bottom-up approach is that there are relatively few free parameters. Throughout the development of the model, we've added additional parameters in those cases where the existing parameters were not sufficiently capturing the behavior, but the over-arching goal was to maintain an understanding of the purpose of each parameter.

4.5.1 Free parameters

4.5.1.1 Learning rate of the model (P1)

The learning rate parameter reflects the rate at which the model adjusts the category representations over the course of the experiment. It can also be thought of as the rate at which a participant in the experiment updates their learning rule or response pattern throughout the experiment. A high value for this parameter evokes large updates to the category representations after feedback.

4.5.1.2 Starting perceptual shaping value (P2)

The perceptual shaping parameter is the same parameter that was discussed in the mathematical models described previously (Chapter 3). This parameter reflects the initial state of the model's ability to discriminate category members in stimulus space (Figure 4.2). Because the model (just like a naïve subject) has no prior knowledge of the category structure, this parameter is initially set at a relatively high value (eg. 10,000). In Figure 4.2, the probability of responding ‘A’ (p(A)) as a function of distance from the boundary is graphed for different values of the perceptual shaping parameter. One consequence of DBT that is shown in this graph is that the greater the distance from the boundary, the more confident the responses (p(A) is closer to 0 or 1). This is true for all values of the perceptual shaping parameter, and as this parameter decreases in magnitude, the category representation is considered to be sharper.

4.5.1.3 Learning rate of the perceptual shaping parameter (P3)

Similar to the learning rate of the model, this parameter determines the rate at which the magnitude of perceptual shaping changes. Since we’re hypothesizing that this is a different process than the one that governs the overall learning of the model, we designate a separate parameter for this learning rate. For example, it’s possible that one could have a high learning rate for the model, but a low learning rate of the perceptual shaping parameter. This would fit a behavioral profile of someone who makes large adjustments in response to feedback, but is slower to incorporate this feedback information into their perceptual processing system.

4.5.1.4 Standard deviation of the Gaussian decision noise distribution (P4)

This parameter was necessary because there was no perceptual noise built into this model. Without some external source of noise, the model would be completely deterministic, thus we added this ‘decision noise’ to the model to emulate human variability in performance. The noise was added to the model by selecting a random number out of a Gaussian distribution with the standard deviation specified by this parameter. A large standard deviation of the Gaussian distribution implies a greater probability of selecting a large random number, which consequently will lead to increased variability in performance.

4.5.2 RB and II system calculations

4.5.2.1 Distance and probability calculations

The model receives input in the form of the frequency and orientation (or x- and y-coordinates in stimulus space) of a sine wave grating stimulus. The output is a categorical judgment as to whether the stimulus belongs in category 'A' or 'B'. Just like a naïve subject, at the beginning of the experiment, the model has no initial concept of the underlying category structure. At the beginning of the first run, a starting position for both an RB and II decision boundary is randomly generated by selecting a number within the range of the x- and y-coordinates, respectively.

On each trial, both the RB and II systems separately calculate the probability that the input stimulus is an 'A' (and by default the probability that the stimulus is a 'B'). This calculation incorporates both the distance of the current stimulus from the decision boundary and the perceptual shaping parameter, both of which get updated on each trial. The series of calculations that result in the estimation of the probability that a stimulus is in category A (p(A)) begin first with the calculation of the mathematical distance from the input stimulus to the decision bound (Equation 4.1). The distance value ($dist_{RB}$ or $dist_{II}$) and perceptual shaping (PS) value are then utilized in the probability calculation (Equation 4.2) (Ashby, 1992). The resulting probability (prob) is equal to the probability that the stimulus was in category A. The probability that the stimulus is in category B (p(B)) is equal to 1-p(A).

Equation 4.1: distance calculations

$dist_{RB} = x_i - RBbound,$

where RBbound=current vertical RB decision bound, x_i=current x-coordinate

$\text{dist}_{II} = (\sqrt{2}/2)*(x_i - y_i + \text{int})$,

where x_i=current x-coordinate, y_i=current y-coordinate, int=current y-intercept of II bound

Equation 4.2: probability calculation

```
ztemp = -(dist/(√(PS)))
zneg = 0
if ztemp < 0:
    zneg = 1
if zneg = = 1:
    ztemp = -ztemp

tt = (-.5*(ztemp)*(ztemp))

if tt < -50:
    tt = -50

fx = (1/√(2*π))*($e^{tt}$)

t = 1/(1+(.33267*ztemp))
prob1 = (0.4361836*t)+(-.1201676*t*t)+(.937298*t*t*t)
```

```
if zneg = = 0:
prob = 1 - (fx*prob1)
elif zneg = = 1:
prob = (fx*prob1)
```

This probability calculation effectively gauges the psychological distance of the stimulus from the decision bound scaled by the perceptual shaping parameter. The RB system, as in the mathematical models, consists of a vertical line in stimulus space that can vary in one dimension (frequency here). The decision boundary in the II system is also a linear bound with a slope fixed at 1.0 that is allowed to vary in its y-intercept.

4.5.2.2 Decision module: selecting a system

The confidence of each system is used to adjudicate between the systems. Here, the confidence of a system is equal to the maximum value of p(A) or p(B). That is, if p(A) > p(B), confidence is equal to p(A) and vice versa. We then calculate the odds that either system would be chosen based on the RB and II system confidence values (Equation 4.3).

Equation 4.3: Odds calculation

$$\mathrm{Odds}_{RB} = \mathrm{con}_{RB}/(1.0 - \mathrm{conf}_{RB})$$

$$\mathrm{Odds}_{II}=\mathrm{conf}_{II}/(1.0 - \mathrm{conf}_{II})$$

Since the model is designed to have competitive RB and II systems, the more confident system should always be selected. However, because we are modeling human behavior, the model needs to have the capability of making the less confident (low probability) choice. This source of noise takes the form of a Gaussian distribution of a variable standard deviation (see P4 above). If the random number selected from the Gaussian distribution with the standard deviation specified by P4 is less than $Odds_{RB}$ - $Odds_{II}$, the RB system is selected. If the random number is greater than the difference of the odds, the II system is selected. The rationale for this system assignment is that a large random number is fairly unlikely to be selected, so only rarely will the system selection deviate from the most confident choice. However, on occasion, when the random number is large enough to overcome the difference of the odds values, the less confident system will be selected. This is an important aspect of the model that allows for strategy shifting. A deterministic model would not be able to account for human response patterns that are susceptible to internal and external sources of noise. By introducing noise into the model we hope to simulate this human-like property.

Another component to the decision module that attempts to account for human behavior is the ability to engage in probability matching. Even if a person is highly confident that a particular stimulus is an A, every so often they will still call that item a B. This type of inconsistent response profile can be modeled by allowing the selected system to make the low probability response a fraction of the time in proportion to the probability of that trial. For example, if the probability that a stimulus is an A is 90%, 90% of the time that stimulus will be called an A, but 10% of the time it will be called a B.

4.5.2.3 Feedback: evaluation of the selected system's response

After either the RB or II system is selected, its chosen response is evaluated and feedback returns to the same system that made the response. If the feedback is positive (system response was correct), no changes are made to the internal representation of the decision bound, but the perceptual shaping parameter is updated (decreased according to its learning rate). The system that was not selected retains the same decision bound and parameter set as the previous trial. Negative feedback leads to the system updating its decision bound in a specific way depending on the location of the incorrect response. For example, if the RB system incorrectly called an 'A' a 'B', this indicates that its decision bound is located too close to the A stimulus space. The new decision bound shifts in the rightward direction in proportion to the learning rate of the model proportional to the distance of the error. In addition, the perceptual shaping parameter increases in response to negative feedback to reflect the poor representation of categories in stimulus space. An identical updating mechanism is used when the II system responds incorrectly.

The feedback evaluation is complicated slightly when the model relied upon probability matching to make its response. That is, if the stimulus was most probably an 'A', yet the model incorrectly responded 'B', the internal representation that led to the original 'A' evaluation should not receive negative feedback. If you consider the analogous thought process going on in a subject, if one is exploring the space by making a response contrary to expectation and this choice leads to negative feedback, this should not change the original hypothesis. On these types of trials, the model receives positive feedback that endorses the high probability choice. The opposite situation could also occur, where the model utilized probability matching and the low-probability response was correct. This indicates that the original high-probability representation needs to be updated because without probability matching it would have selected an incorrect response. Thus on these sort of trials, the model receives negative feedback despite the fact that the actual system response was correct.

4.6 Model fitting

Evaluating the performance of PINNACLE in fitting the available behavioral data was a multi-step process. The first step was to verify that the model produced learning curves similar to the human subjects in the experiment. This meant matching the model's accuracy across 4 80-trial blocks to the RB and II group accuracy (in both Experiments 1 and 2).

The next step was to use the model to make predictions of individual subject performance on a trial-by-trial basis. This step required that the model find the optimal set of parameters that generate a behavioral profile that most closely matches each subject. Importantly, the model does not 'know' whether a subject is receiving an RB or II category structure. It merely finds the best fitting parameters that can account for the set of category decisions the person happened to make.

4.6.1 Verifying PINNACLE behavior matches average group data

This first evaluation of PINNACLE can be conceptualized as demonstrating that the model structure produces learning that resembles human performance. Practically, this means verifying that the model is capable of generating accuracy levels across subject groups that approximate those of the actual RB and II subject groups in both Experiment 1 and 2.

To simulate the RB and II experimental groups, PINNACLE received both RB and II category structures in the same trial orders as the human subjects. On each trial, the model generated a prediction of category membership based on its internal representation of the category structure (described previously). After all subjects had been simulated, the average accuracy and standard deviation per block of the RB and II model-simulated groups was compared to these actual subject values observed in the experiment.

The 'fit value' of a single pass through the stimuli was based on the sum of squares difference between the model-based accuracy and standard deviation and the subject accuracy and standard deviation. Depending on this fit value, the search algorithm adjusted the parameters and repeated the entire process working to minimize this value. The results of this minimization procedure identified the optimal parameter set that produced performance that best accounted for the actual subject group behavior. The relative importance of fitting the accuracy levels over that of fitting the standard deviations meant that we weighted the fit calculation in favor of the accuracies (Equation 4.4).

Equation 4.4: Group model fit calculation

$$\text{RB acc fit} = \sum(\text{RBmodacc}_n - \text{RBsubjacc}_n)^2 \text{ where n = 1 through 4}$$

$$\text{II acc fit} = \sum(\text{IImodacc}_n - \text{IIsubjacc}_n)^2 \text{ where n = 1 through 4}$$

$$\text{RB st dev fit} = \sum(\text{RBmodstdev}_n - \text{RBsubjstdev}_n)^2 \text{ where n = 1 through 4}$$

$$\text{II st dev fit} = \sum(\text{IImodstdev}_n - \text{IIsubjstdev}_n)^2 \text{ where n = 1 through 4}$$

$$\text{Fit value} = [(\text{RB acc fit}) + (\text{II acc fit})] + [(\text{RB st dev fit}) + (\text{II st dev fit})]/1000$$

Since PINNACLE is intended to be a model of categorization behavior, the particular stimulus sets or trial orders should not affect the operation of the model. Rather than fit the data from Experiment 1 and 2 separately, we chose to fit the model to the combined dataset of both experiments together. Not only does this provide a variety of trial orders, but also stimuli from different parts of perceptual space. The resulting best fitting parameters should thus have the most accurate correspondence with reality.

4.6.1.1 Amoeba simplex search algorithm

The search algorithm used to find the optimal set of parameters was an amoeba simplex search method (Press et al., 1992). Based on the number of starting parameters (N), amoeba generates a 'simplex' which is a list of N+ 1 starting parameter sets. The model generates a fit value for each of these starting parameter sets, and the direction of the search within parameter space is determined by the lowest of these fit values. As amoeba iterates through the parameter sets, it changes the parameters in such a way as to move in the direction of the lowest fit values. A benefit to this search algorithm is that it naturally gravitates toward low regions of parameter space, making the search faster and more efficient than other so-called 'brute force' approaches.

4.6.1.2 Group data fits: 4 parameter model

Using the amoeba simplex downhill search algorithm, we identified the set of 4 parameters that minimized the difference between the group averages and the model outputs. These model fits were evaluated for 12 different feedback models as a way of testing assumptions about feedback processing (see section 4.6.2.2). However, with only 4 parameters, the search algorithm could not find a good fit to the data (Figure 4.3) in any of the different feedback models.

4.6.1.3 Group data fits: 5 parameter model

One of the assumptions with the 4-parameter model is that there is a single learning rate for the model, which applies similarly to the RB and II systems. In light of the hypothesized processing differences between these systems and the inability of the 4-parameter model to find a good fit to the data, we surmised that the systems may require independent learning rates. Thus, we added a fifth parameter to the model so now the algorithm was searching for the best fit with an RB learning rate, II learning rate, learning rate of the perceptual shaping parameter, initial perceptual shaping value and standard deviation of the decision noise.

When separate RB and II learning rates were incorporated into the model, the 5-parameter model returned a much lower optimal fit value than with the 4-parameter model (Figure 4.4). This suggested that rather than having a single learning rate, learning in the RB and II systems may require independent processing capabilities.

4.6.2 Model assumptions

When constructing a model of human behavior, a number of practical issues inevitably arise that can only be resolved by making certain assumptions. Being aware of these assumptions is important as they can guide the interpretation of the results of the model behavior as well as provide points for improvement in future versions. The operation of PINNACLE relies upon a number of important initial assumptions, which are discussed briefly here and are laid out in Table 4.1.

4.6.2.1 Testing feedback processing and system interaction in PINNACLE

Constructing a first version of PINNACLE meant making certain assumptions, and testing these assumptions is an important step in increasing our confidence that the model is capturing some cognitive reality. A number of these initial assumptions are related to the feedback processing of the model as well as the system dynamics. By comparing different versions of PINNACLE that instantiate different mechanisms of feedback and system interaction, we can determine the model that best accounts for the data. These feedback models fall into two general categories that are described below.

4.6.2.1.1 RB and II system interaction

The goal of PINNACLE is to model two interactive learning systems (RB and II), but this does not rule out the possibility that there are more than two systems in the brain available to acquire categorical information. By narrowing the range of possibilities, we hope to understand the underlying mechanisms of these two learning systems, and as the model becomes more specific and complex, additional capabilities can be added. For example, a number of pieces of evidence support the idea that there is a posterior circuit involving sensory cortex that is important for perceptual learning processes in categorization (Reber et al., 2003). The current version of PINNACLE does not attempt to account for this third 'category learning system'.

The nature of the interaction between the RB and II systems is assumed to be competitive, in that the more successful system will maintain an advantage over the less successful system. This assumption was based on evidence from animal work as well as neuroimaging of humans in different learning situations, but the evidence does not definitively support the competitive systems hypothesis. As discussed in Chapter 1, there are several other types of interaction that are also plausible: cooperative interactive systems, parallel independent systems and a combination of different modes of interaction. Any of these described methods of system interaction are potentially relevant to our model of category learning. To address these possibilities, additional versions of the model that manipulate the interactive relationship of the RB and II systems are competitively assessed. The degree to which the model fits the data guides us toward the type of interaction that most closely reflects reality in the context of the model's framework.

To test the idea that the systems learn simultaneously in parallel, feedback is allowed to return to both systems, regardless of the system that made the response (see Model 2 in Table 4.2). This feedback will potentially influence the systems differently according to the prediction of the system. The most confident system will still dictate the overt response, but each system will be allowed to update accordingly. The appropriate system for the task should still be the one that is most likely responsible for the category decisions, but the difference from this and the competitive model is that the alternate system would simultaneously develop an appropriate category representation.

Other variations of this idea are captured in Models 4 and 5 in Table 4.2. In Model 4, the RB system receives feedback regardless of the system that was active on that trial, but the II system only receives feedback on trials where it was active. The converse of this mechanism is represented in Model 5.

4.6.2.1.2 Positive vs. negative feedback

In learning, feedback is a critical feature that helps shape behavior. Feedback generally comes in two varieties: positive and negative. Positive feedback endorses the most recent behavioral response and negative feedback indicates that an error has occurred. In the context of the PINNACLE model, learning in the RB and II systems is assumed to occur similarly with the incorporation of negative feedback. It is only on these erroneous trials that the category representation is updated. On positive feedback trials the current category representation is endorsed such that the noise associated with the decision bound is reduced, but no changes are made to the representation itself. There is some evidence, however, that positive and negative feedback differentially affect learning depending on the task demands.

In a set of early rule-based category learning tasks, participants faced with a two-choice response task exhibited more effective learning with negative rather than positive feedback (Meyer and Offenbach, 1962). In this case, negative feedback was considered to be more informative than positive feedback because it both signaled the participant that his or her hypothesis was incorrect and also signaled which was the correct response. As discussed previously (Chapter 1), RB learning progresses even in the absence of any type of feedback, which has not been the case for II learning (Ashby et al., 2002). In one study of II category learning, the presence of feedback was deemed necessary for learning to occur, but they did not differentiate between positive and negative feedback (Ashby et al., 1999). More recently, a study of II category learning under different feedback conditions found that both positive and negative feedback signals are critical for learning (Ashby and O'Brien, 2007). In this study, participants were asked to learn II category structures either with only positive feedback, only negative feedback, or full-feedback (both positive and negative). Model-based analyses of these different groups revealed that positive or negative feedback alone leads to inappropriate RB strategy application and only the full-feedback condition fostered II strategy use.

Based on these results, one might predict that positive and negative feedback should differentially mediate learning in the RB and II systems of PINNACLE. Again using different versions of PINNACLE that manipulate the feedback processing within and between the RB and II systems can address these questions. Model 6 (Table 4.2) instantiates the hypothesis that the RB system only receives negative feedback whereas the II system receives both positive and negative feedback. In contrast, Model 7 assumes that the II system only receives negative feedback and the RB system receives both kinds of feedback information.

In Model 1, both the RB and II system representations are updated similarly in response to negative feedback (the standard model). Model 3 instantiates the idea that it is only negative feedback that changes the category representation whereas Model 12 instantiates the idea that learning only occurs with positive feedback. Models 10 and 11 are similar to the aforementioned models, but they also restrict the feedback information to the system that was assigned to that particular trial. In Models 8 and 9, the idea that the RB and II systems benefit from different types of feedback is instantiated. In Model 8, if the RB system is selected, it will only receive positive feedback whereas if the II system is selected it will only receive negative feedback. Model 9 represents the opposite situation where the RB system receives negative feedback and the II system positive feedback.

4.6.2.2 Feedback models

A verbal description of each version of the model is laid out in table 4.2. The initial version of the PINNACLE model is described under the label ‘Model 1’. The other 11 model versions contain different hypotheses about system interaction and feedback as described above. To determine the optimal feedback/system interaction model, we fit the available data with each of these models. As an additional test, these models were tested both with the 4- and 5-parameter versions of PINNACLE.

Out of the 4-parameter model fits, Model 6 returned the minimum fit value (Table 4.3A). Out of the 5-parameter model fits, Model 2 returned the minimum fit value (Table 4.3B). A comparison of the 4- and 5- parameter models revealed that the 5-parameter models are superior on the whole. That is, for each version of the model, the 5-parameter version returned a lower fit than the 4-parameter version. Since Model 2 returned the minimum fit value out of all the different model versions, the hypothesis that the RB and II systems are competitive may not be appropriate (although Model 1 returned the second lowest fit value).

From this competitive feedback modeling and group data fitting come three important results. The first is that the 5-parameter version of PINNACLE is superior to the 4-parameter version. The best account of the behavioral data occurs with a model that is capable of applying distinct learning rates to the RB and II systems. The second is that no matter which system was assigned to a particular trial, feedback should return to both systems. This interaction is more in line with parallel processing ideas discussed in Chapter 1. The third result is the identification of the best fitting Gaussian sigma parameter value for the 5-parameter model using feedback model #2. This parameter value (1.204) defines the standard deviation of the noise distribution for the individual subject fits discussed in section 4.6.3.

4.6.3 Model fitting to individual subject data

The next step after determining that PINNACLE successfully accounted for the average group data was to use the model to fit individual subject behavior. This step represented one important goal of the model, which was to identify RB and II system engagement on a trial-by-trial basis. That is, rather than testing whether the model could match human behavior when presented with the real category structures, we minimized the model fit to the subject's actual responses. This meant that overall accuracy was no longer the important measure, rather it was the minimization of the difference between the model behavior and actual subject behavior (good or bad).

As with the group model fits, on each trial both the RB and II systems generated p(A) given the stimulus input and the current state of the RB and II decision boundaries. Based on the subject's actual response (A or B), the system that was more likely to have generated that same response was assigned to that particular trial. If both systems generated the same response (i.e. both systems said 'A'), PINNACLE selected the one that had higher confidence. Since the actual subject responses were used here, the decision noise parameter was no longer necessary, but the confidence prediction was still an important component of the systems. Feedback returned to both systems since this was the best fitting feedback version in the group data fitting procedure. The idea here, again, was that we were trying to find the best set of trial-by-trial model outputs that account for the subjects' responses.

To identify the set of parameters that best account for the subject's data, we used the Maximum Likelihood Estimation (Ashby, 1992). The method of evaluation of the fit value for a particular set of parameters for an individual subject was slightly different from the previously described fit value calculation for the group data. Instead of calculating the sum of squares difference between the model block averages and the subject block averages, on each trial we calculate the odds of the model producing the same response as the subject. If the model with that particular set of input parameters was very unlikely to have produced the same response as the subject, that trial contributes a bad fit value (high number) to the total fit. This process is repeated throughout the entire 320 trials. The fit value reflects the sum of the odds on each trial, and the goal of the search algorithm is to minimize this value.

The output of the model is a vector that contains a prediction on each trial as to whether the RB or II system was active in producing the behavioral response of the subject. An example of an RB and II subject profile is shown in Figure 4.5. In general, the RB system is more confident over time (produces values that are closer to 0 or 1) than the II system in the RB subject and vice versa for the II subject. These trial-by-trial predictions are utilized in the fMRI data analyses discussed in Chapter 5.

4.6.3.1 Simulation 1: Experiment 1 data

The PINNACLE model was used to fit the behavioral data from Experiment 1. To evaluate the overall pattern of results, we aggregated the data over the four blocks for each group of subjects. In Figure 4.7A, the graph shows the average number of RB and II trials in the RB subject group within each of the 4 blocks. Figure 4.7B shows a similar graph for the average number of model identified trials in the II groups. The individual subject characterization across blocks is also shown in Figure 4.6B. For comparison purposes, the schematic for the mathematical models is shown in Figure 4.6A. The averaged trial-by-trial predictions are highly similar to the mathematical model characterization of the data on a block-level. This comparison reveals that PINNACLE is appropriately identifying RB and II strategy-use in a way that is consistent with the simpler DBT models.

4.6.3.2 Simulation 2: Experiment 2 data

PINNACLE was used to fit the behavioral data from Experiment 2. To evaluate the overall pattern of results, we aggregated the data over the four blocks for each group of subjects. In Figure 4.9A, the graph shows the average number of RB and II trials in the RB subject group within each of the 4 blocks. Figure 4.9B shows a similar graph for the average number of model identified trials in the II groups. The individual subject characterization across blocks is also shown in Figure 4.8B. For comparison purposes, the schematic for the mathematical models is shown in Figure 4.8A. As in Experiment 1, there was a high degree of similarity between the

DBT-based mathematical model fits and PINNACLE's averaged trial-by-trial predictions over an 80-trial block. Also similar to the mathematical model results, PINNACLE revealed a greater amount of strategy-switching within the RB and II subject groups than observed in Experiment 1. This was true particularly in the II group who were frequently utilizing an inappropriate RB strategy.

4.7 Discussion

PINNACLE represents the implementation of the hypothesized cognitive processes underlying RB and II category learning systems. This cognitive model builds upon the results from the mathematical characterization of behavior over 80 trial blocks. Shifting from static DBT models to a dynamic trial-by-trial model required a number of important changes that ultimately better reflect our hypotheses of the operation of these systems in the brain. Namely, a healthy human has the capacity to utilize both the RB and II systems on any given learning task. By allowing these systems to interact within PINNACLE, we hoped to emulate the processing that occurs during the acquisition of these sine-wave stimuli categories.

Simulations of the behavioral data from fMRI experiments 1 and 2 provided PINNACLE with a variety of performance profiles as well as stimulus distributions to work with. By comparing the minimum fit value of different versions of PINNACLE, we could test different assumptions about the model's operation. As a result of this process, a number of aspects of the model changed, including the number of free parameters, the feedback structure and the type of system interaction. Each step in the evaluation process contributed toward the ultimate goal of developing a model that was capable of trial-by-trial predictions of RB and II strategy-use in individual subjects.

The subject fits using the DBT-style mathematical models (Chapter 3) serve as a point of comparison to PINNACLE's predictions. In general we expected to see that the block-level strategy use was preserved between the two styles of models. That is, the dominant strategy utilized in a block of 80-trials in PINNACLE should largely agree with the static DBT assessment of strategy-use. Inspection of PINNACLE's block-level predictions confirms that the course assessment of strategy-use is consistent between the models in both Experiment 1 (Figure 4.6) and Experiment 2 (Figure 4.8).

The more detailed trial-by-trial predictions of system engagement become important when considering the event-related fMRI data. PINNACLE provides an alternative to more traditional fMRI analysis techniques, and the incorporation of functional activity allows us to test more specific hypotheses about the structure of the model. Chapter 5 discusses this idea in detail, but similar to the process of model fitting, assessments of the model behavior with fMRI data allows us to examine the assumptions we've made in building this model.

CHAPTER 5: TESTING PINNACLE PREDICTIONS WITH FMRI DATA

5.1 Overview

Functional neuroimaging acts as an effective bridge to connect sophisticated cognitive models with the operation of the underlying neural structures. In the section on mathematical models (Chapter 3), model-guided fMRI data analysis revealed several additional components of the RB and II category learning networks. As PINNACLE is a more sophisticated model that accounts for the trial by trial interaction between these systems, PINNACLE's predictions of behavior should facilitate a cleaner separation between RB and II trial types. Applying a similar trial-sorting method to the fMRI data enhances the specificity of the contrasts that rely on this separation. Further, the incorporation of fMRI data can provide opportunities to test the underlying assumptions guiding the model.

5.2 PINNACLE model fitting enhances fMRI data analysis

In traditional data analysis methodology, behavioral criteria such as accuracy or stimulus type are typically used to segregate trial activity in event-related fMRI paradigms. PINNACLE provides us with another method by which to group trial activity that is based on the model's prediction of which system was active on any given trial. The data discussed here demonstrates the effectiveness of this PINNACLE-based trial sorting in identifying regions of activity associated with RB and II category learning.

5.2.1 Methods

The PINNACLE model of category learning was used to fit each subject's behavior in both fMRI experiments in order to estimate strategy use on a single trial basis (see Chapter 4).

The resulting vectors were used to organize the data in a way that isolated activity during RB trials from that during II trials. The comparison of correct to incorrect categorization (as defined in Chapter 2 as *successful categorization*) was also used here to identify regions of activity associated with RB and II learning.

5.2.2 fMRI Experiment 1

5.2.2.1 Results

5.2.2.1.1 Behavioral observations

One initial assessment of PINNACLE is the accuracy of the RB and II subject groups on both RB and II trials. In general, one would expect that the correct trials for RB subjects should coincide with assignment to the RB system and similarly, correct trials for II subjects should coincide with assignment to the II system. For the RB group, across all runs the accuracy on RB trials was 84.3% (SE=0.019) and on II trials was 39.8% (SE=0.031). For the II group, across all runs the accuracy on RB trials was 47.9% (SE=0.03) and on II trials was 89.2% (SE=0.014). The accuracy per run for both the RB and II groups is shown in Figure 5.1.

5.2.2.1.1 Whole-brain activity

5.2.2.1.1.1 All blocks

On trials where PINNACLE predicted the RB system was active, a contrast of correct and incorrect activity within the RB subject group revealed activity in the anterior cingulate, parietal and frontal cortices (Figure 5.2A). A similar contrast in the II subject group demonstrated strong bilateral caudate activity (Figure 5.2B). On trials where PINNACLE predicted the II system was active, in RB subjects the left anterior MTL was more active for correct than incorrect trials (Figure 5.3A). In II subjects, the left posterior body of the caudate as well as posterior cingulate and occipital cortex were more active for correct than incorrect trials (Figure 5.3B).

5.2.2.1.1.2 Best fitting blocks

When correct trial activity was restricted only to those blocks that were best fit by PINNACLE (Figure 5.4), we observed significant regions of activation in right PFC greater for RB than II trials and in right posterior visual association cortex greater for II than RB trials. This pattern of activity is highly similar to that observed in the DBT-based block sorting in Figure 3.2. This observation supports the effectiveness of PINNACLE-based trial sorting in isolating additional regions of the RB and II networks not found with traditional fMRI data analysis methods.

5.2.2.1.2 Region-of interest analysis

Similar contrasts to those done in the whole-brain were repeated within the MTL and caudate ROIs to determine whether PINNACLE-based trial sorting could identify activity differences not observed in less sensitive subtractions. Similar to the original fMRI analysis that did not include model predictions (Figure 2.3), within the MTL ROI (Figure 5.5), successful categorization activity during RB trials elicited activity in the left anterior MTL for RB subjects. The II subjects also showed a successful categorization difference bilaterally in the MTL ROI (Figure 5.5) during RB trials. During II trials, bilateral anterior MTL activity was observed for RB subjects and left anterior MTL activity for II subjects. It is worth noting that the strongest activity occurred during RB trials, regardless of the subject grouping.

In the caudate ROI (Figure 5.6), successful categorization activity was observed during RB trials in the bilateral head of the caudate in the RB subject group and the bilateral head and body of the caudate in the II subject group. During II trials, the RB subjects had no significant regions of activity in the caudate, but the II subjects had successful categorization activity in the bilateral head and body of the caudate.

5.2.2.2 Discussion

The first goal of PINNACLE-based trial sorting was to identify additional components of the RB and II category learning neural networks. The best analysis to address this goal was to restrict the trial activity to only that which corresponds to the active system. That is, to visualize the RB categorization network, within the RB subject group, isolate activity associated with trials that PINNACLE designated as corresponding to the RB system. Then, within these trials, contrasting the correct and incorrect trial activity should illuminate the RB and II network activity. The previous whole-brain analysis of RB and II network activation that did not depend on PINNACLE's assessment of system involvement on each trial (see Chapter 2) can serve as a comparison here.

In the RB subjects in the first fMRI experiment, PINNACLE-based trial sorting revealed a network of regions throughout the brain that are active during successful RB category learning. The anterior cingulate cortex, left medial frontal gyrus and cuneus were more active for correct than incorrect RB trials. This same network of regions was also active in the original fMRI analysis (Figure 2.2A) with the noteworthy absence of the MTL activation (discussed below).

The only II categorization activity in the original fMRI analysis (Figure 2.2B) in the II subject group was observed in the body of the caudate. As shown in Figure 5.3B, PINNACLE-identified II trials in II subjects isolated a corresponding region in left posterior body of the caudate that was more active for correct than incorrect II trials. Two additional regions of activity were observed in the posterior cingulate cortex and visual association cortex. These regions were not observed to be active in the original II subject successful categorization network, yet the COVIS theory of category learning predicts the presence of posterior visual regions in association with caudate activity.

A more restricted analysis of only the best fitting blocks of 80 trials identified a number of regions of activity similar to those found with DBT-based modeling. Within the blocks that were best fit with PINNACLE, we contrasted activity associated with correct RB trials and correct II trials. That is, trials in which PINNACLE predicted the RB system was successfully engaged in learning were compared to trials where the II system was successfully learning. A region in the right PFC was more active during correct RB than II trials. The opposite effect was observed in right posterior visual association cortex where activity was greater during correct II than RB trials. These regions were also identified in the DBT model-based analysis of the best fitting blocks of RB and II strategy use and are consistent with initial hypotheses of the involvement of these regions in RB and II learning.

In addition to examining the RB and II group activity during RB and II trials, the inverse of this analysis has the potential to identify activity associated with the non-optimal strategy. Simply, what regions in RB subjects were more active during successful II category learning and what regions in II subjects were more active during successful RB category learning?

Given that we are studying healthy human behavior here, and PINNACLE presumes that there are 2 interactive category learning systems in the brain, it is tempting to interpret the RB and II trial activity within each subject group (or as a larger single group) as representing the network of RB and II category learning systems. However, one of the limitations of PINNACLE is its tendency to assign incorrect responses to the non-dominant system. The accuracy of PINNACLE's trial-by-trial predictions is summarized in Figure 5.1, and demonstrates the point that the majority of incorrect responses in RB subjects are attributed to the II system and the incorrect II subject responses are attributed to the RB system. Thus, a caveat to interpreting the non-dominant system activity is that it may be weighted by incorrect trial activity.

An examination of RB trial activity in II subjects shows strong bilateral body and head of the caudate activation. In these II subjects, PINNACLE has indicated that the RB system is engaged, yet we do not see the 'typical' RB categorization regions. Similarly, II trial activity in RB subjects does not reveal the typical II categorization network. Instead, the left anterior MTL is more active for correct than incorrect II trials. The distribution of incorrect trials may provide a potential explanation for this surprising result. Since the successful categorization contrast depends on the difference between correct and incorrect trials and the non-dominant system contains the bulk of the incorrect trials, the resulting regions of activity are more consistent with the dominant system. For example, the MTL activity found in the RB subjects during the supposed II trials may simply reflect the fact that the MTL was more engaged (here, more deactivated) during the incorrect II trials, but these trials were mis-labeled by PINNACLE. Similarly, the lack of MTL activity in the RB subjects during RB trials may be due to the lack of incorrect RB trials that were designated as incorrect II trials.

Given the important role of both the MTL and caudate in RB and II category learning, the same PINNACLE-based fMRI analyses as were done in the whole brain were done in the MTL and caudate ROIs. In the MTL, the hypothesis was that RB trial activity would be greater than that of II trial activity in both the RB and II subject groups. This pattern was generally observed, although there was also evidence of II trial activity in RB subjects (Figure 5.3, top right panel). As discussed previously, these clusters of MTL activity associated with II trials in RB subjects may be attributable to the inappropriate assignment of incorrect trials to the II system.

In the caudate ROI, as predicted, in II subjects the body is active bilaterally during II trials. This is also true during RB trials, but is not true in RB subjects during RB trials. Rather, the head of the caudate is more active during RB trials in both RB and II subjects. The head of the caudate is thought to be involved in hypothesis testing and switching amongst strategies, so this pattern of activity could support this hypothesis.

5.2.3 fMRI Experiment 2

5.2.3.1 Results

5.2.3.1.1 Behavioral observations

For all subjects, accuracy rates were aggregated across RB and II trials according to PINNACLE's predictions. For the RB group, across all runs the accuracy on RB trials was 85.7% (SE=0.02) and on II trials was 53.2% (SE=0.03). For the II group, across all runs the accuracy on RB trials was 54.3% (SE=0.015) and on II trials was 78.9% (SE=0.015). The accuracy per run for both the RB and II groups is shown in Figure 5.7.

The increased variance in the stimuli in Experiment 2 also allowed for an analysis of trials based on perceptual distance from the category boundary (Figure 5.8). On RB trials the close trial accuracy was 65.7% (SE=0.019) and the far trial accuracy was 77.4% (SE=0.031). On

II trials the close trial accuracy was 61.6% (SE=0.014) and the far trial accuracy was 77.6% (SE=0.026). There was a significant interaction between the close and far trial accuracy between the RB and II trials (F=4.916, $p<0.05$) in RB and II subjects.

5.2.3.1.2 Whole-brain activity

5.2.3.1.2.1 All blocks

On trials where PINNACLE predicted the RB system was active, a contrast of correct and incorrect activity within the RB subject group revealed activity in the left frontal and parietal cortex greater for correct than incorrect trials and anterior cingulate, insula and cuneus greater for incorrect than correct trials (Figure 5.9A). A similar contrast in the II subject group identified activity in the left caudate body and occipital cortex greater for correct than incorrect trials and bilateral insula and anterior cingulate greater for incorrect than correct trials (Figure 5.9B). On trials where PINNACLE predicted the II system was active, in RB subjects the frontal cortex, anterior cingulate and cuneus were more active during incorrect than correct trials (Figure 5.10A). In II subjects, the insula and frontal cortices were more active during incorrect than correct trials (Figure 5.10B).

5.2.3.1.2.2 Best fitting blocks

When trial activity was restricted only to those blocks that were best fit by PINNACLE (Figure 5.11), we observed significant regions of activation in medial PFC greater for RB than II trials and in right posterior visual association cortex greater for II than RB trials. Only the correct trials were used in this contrast as there were relatively few incorrect trials in the best fitting blocks. This pattern of activity is highly similar to that observed in the DBT-based block sorting in Figure 3.2. It further supports the effectiveness of PINNACLE-based trial sorting in isolating additional regions of the RB and II networks not found with traditional fMRI data analysis methods.

5.2.3.2.2 Region-of interest analysis

Similar contrasts to those done in the whole-brain were repeated within the MTL and caudate ROIs to determine whether PINNACLE-based trial sorting could identify activation differences not observed in less sensitive subtractions. Within the MTL ROI (Figure 5.12), successful categorization activity during RB trials elicited activity in bilateral anterior MTL for RB and II subjects. During II trials, no activity was observed for RB subjects, but II subjects showed left anterior MTL activity.

In the caudate ROI (Figure 5.13), successful categorization activity was observed during RB trials in the bilateral head and body of the caudate in II subjects, but no significant clusters of activity were observed in the RB subjects. During II trials, both RB and II subjects sustained activity in the right head/body of the caudate greater for incorrect than correct trials. In contrast, only the II subjects had activity in the posterior body of the caudate for correct greater than incorrect trials.

5.2.3.2 Discussion

As with the data from Experiment 1, PINNACLE-based trial sorting was used to identify regions of functional activity associated with RB and II trials. In the RB subject group, the cuneus and left IFG were more active for correct than incorrect RB trials. Bilateral insula and ACC were more active for incorrect than correct RB trials in the RB subject group. A comparison to the original fMRI analysis (Figure 2.8A) shows that the insular regions were similarly identified as eliciting greater activity during incorrect than correct trials. The difference in stimuli and degree of difficulty between Experiment 1 and 2 may have contributed toward the greater evoked activity for incorrect than correct trials in the RB subjects.

The II subject group also displayed primarily greater activity for incorrect than correct trials, both during RB and II trials. In addition to bilateral insular regions, the right medial frontal gyrus was more active for incorrect than correct II trials. This is the same region as that found in the non-model based analysis (Figure 2.8 B). There were several small regions identified in posterior caudate that exhibited successful categorization activity during II trials, but not to the same extent as seen in the original analysis.

Within the II subjects, an examination of the non-dominant system activity reveals a similar phenomenon to that seen in Experiment 1. Namely, the regions of activity found in the II subjects during PINNACLE-identified RB trials are consistent with typical II categorization areas. The body of the caudate and regions in the posterior visual association cortex are more active for correct than incorrect RB trials in II subjects. Again, this paradoxical effect may represent PINNACLE mis-labeling the incorrect II trials as RB trials.

While the PINNACLE-identified RB and II trial analysis did identify a number of regions of activity not observed in the accuracy-based fMRI analysis, restricting the trial activity to only those blocks that were best fit with PINNACLE allowed for a more sensitive contrast of activity. Specifically, the best examples of RB and II strategy use across an 80-trial block isolated functional activity in regions consistent with our initial hypothesized RB and II networks. Activation in the PFC and posterior visual association cortex was also observed in Experiment 1 using a similar restricted analysis, further supporting the important roles of these regions in RB and II learning, respectively.

The activity observed in the MTL ROI is very similar to that seen in Experiment 1. Both the RB and II subjects demonstrated MTL activity during RB trials yet this activity was greater in the RB subjects. There was a small region in left anterior MTL in the II subjects during II trials, which again may be attributable to these incorrect RB trials being mis-labeled. In the RB subjects, however, there was no evidence of MTL activity during the II trials either in the whole brain or the anatomical ROI. The aforementioned performance difference between the RB subjects in Experiment 1 and 2 may contribute to this difference in the MTL activity. Because there were relatively few incorrect trials in the RB subject group in Experiment 2, it is possible that they did not provide a sufficient contrast to the more abundant correct trials.

Within the caudate ROI, the II trials elicited significantly more activity during correct than incorrect trials in II but not RB subjects in posterior caudate. The II subjects also demonstrated a positive successful categorization difference during RB trials throughout the body of the caudate. The only difference within the RB subjects in this ROI was in the right anterior body of the caudate. This region was significantly more active during incorrect than correct II trials both in the RB and II subjects suggesting some common mechanism between these two groups.

5.3 Using PINNACLE-based trial sorting to test model assumptions

An important goal of developing a model such as PINNACLE was not only to enhance existing data analysis techniques, but also to provide the tools with which to challenge the assumptions of the model. The following analyses demonstrate such potential applications of PINNACLE using the existing fMRI data. These examples are the first attempts at addressing the assumptions underlying PINNACLE using a model-based trial sorting technique.

5.3.1 Identifying decision module competition-related fMRI activity

5.3.1.1 Background and motivation

A major component of PINNACLE is the 'decision module' that serves to adjudicate competition between the RB and II systems. We hypothesize that the region of the brain most likely to support this role is the PFC. While the animal and neuroimaging data discussed in Chapter 1 does implicate the PFC in this role, PINNACLE-based trial sorting provides an opportunity to test this assumption.

In the following analysis, PINNACLE-based predictions were used to analyze the behavioral data acquired in Experiments 1 and 2 to identify instances of peak competition between the RB and II systems. The rationale behind this analysis is that the DM should be active on every trial, but *more* active on trials that require a more difficult decision between the systems. The most difficult decision between the RB and II systems should occur when both systems are highly confident. Instances of competition between systems only make up a fraction of the total number of trials (Figure 5.12A). The remainder of the trials can be divided into two types: non-competitive and undefined trials. Trials on which one system is highly confident while the other system is not confident make up the 'non-competitive' trials. The rest of the trials that we are calling 'undefined' are those in which both systems exhibit low confidence, and the corresponding activity may not be particularly informative regarding questions of system interaction.

After identifying a number of these instances within subjects, averaging the functional activity associated with these trials should reveal the neural structures actively involved in this system mediation. Due to the fact that there are relatively few trials that fit this 'competition' profile, it was technically difficult to construct the fMRI contrast in a statistically optimal manner. The best way to address this question would be to design a new study specifically aimed at fostering situations of high competition between systems. Numerous instances of competitive interactions could then contribute toward the fMRI data analysis.

With the current datasets, however, it is still possible to examine this question of competition-related activity. Since the DM is hypothesized to be active on every trial, we can isolate the search space to only those regions that are consistently active to all trial types. A functional ROI based on this 'all events' contrast can then be used as a mask within which we search for competition-related activity. Should PFC activity be correlated with these model-identified instances of competitive interaction, we will have greater confidence that the current structure, which was intended to emulate the PFC and its interconnections with MTL and basal ganglia memory systems, is an appropriate neural framework for this cognitive module.

5.3.1.2 Methods

In all subjects in both Experiment 1 and 2, trials were marked as exhibiting competition (C), non-competition (NC) or undefined interaction (U). In the C trials, both the RB and II system confidence was required to be greater than 75%. In the NC trials, either the RB or II system confidence must have been greater than 75% while the other system must have been less than 75% confident. In the U trials, both system confidence levels must have been less than 75%.

The fMRI data analysis utilized in Experiment 1 and 2 was identical. In both cases, the peak-to-baseline contrast of trial-evoked activity across all trials and all subjects provided the set of voxels within which we utilized PINNACLE-based trial sorting. This smaller volume allowed for greater sensitivity than is afforded when searching the entire brain for trial-evoked activity. The grouped functional ROI was used to mask each individual subject's contrast of competition-related activation. The resulting t-test then used these masked functional datasets to isolate significant clusters of activity throughout the whole brain.

5.3.1.3 Results

In Experiment 1, the average number of C trials per run was 19, NC trials was 28 and U trials was 33 (Figure 5.12A, left). The accuracy on these different trial types is shown in Figure 5.12B (left). On C trials the average accuracy per run was 84.7%, on NC trials was 80.7% and on U trials was 68.25%.

In Experiment 2, the average number of C trials was 21, NC trials was 34 and U trials was 24 (Figure 5.12A, right). The accuracy on these different trial types is shown in Figure 5.12B (right). On C trials the average accuracy per run was 83.2%, on NC trials was 76% and on U trials was 62.4%.

Figure 5.12C shows the contrast of C versus NC trial activity in both Experiment 1 and 2. The right DLPFC and bilateral motor cortex are more active during C than NC trials. In Experiment 1 there is an additional region active in the posterior parietal cortex.

5.3.1.4 Discussion

PINNACLE assumes that the brain has at least 2 category learning systems that are both subject to mediation by the DM. This means that on each trial, both the RB and II systems may develop a prediction of category membership, but only one system is allowed to influence the subsequent motor response. The role of the DM is to adjudicate the assignment of the RB or the II system to a particular trial.

The analysis here was aimed at isolating functional activity during moments where the DM was posed with a particularly difficult challenge. When both the RB and II systems are highly confident in their response it should be more difficult for the DM than a situation where one system is much more confident than the other. As hypothesized, a region in right DLPFC exhibits competition-related activation in both Experiment 1 and 2. The specific action of the DM/DLPFC on these trials may be to actively inhibit one system so that the other system can send activation to the motor system to make the appropriate motor plan. This type of inhibitory role of the DLPFC has been observed using fMRI with task-switching paradigms where one task-response needs to be inhibited to allow the other to progress (Sylvester et al., 2003; Yeung et al., 2006).

5.3.2 Identifying off-system activity with PINNACLE

5.3.2.1 Background and motivation

On each trial, PINNACLE assigns either the RB or II system to the behavioral response. However, in the model both systems generate an independent prediction. There is a confidence associated with each system's best guess for the behavioral response given the current internal decision boundary. In the previously described analysis of RB and II trial activity (section 5.1), only the dominant system (or 'on-system') is used to define the functional activity. Given that the non-dominant system (or 'off-system') receives feedback under the current feedback mechanism, this analysis is aimed at examining the incorporation of the feedback in the off-system. It is important to note here that the feedback may have different effects on the on- and off-systems. Even if the on-system was incorrect, if the off-system made the correct prediction, the system does not treat the feedback as negative. It will endorse its current system representation as if it had received correct feedback. Similarly, if the off-system made an incorrect prediction it will update its representation accordingly even if the feedback to the on-system was correct. For this reason, an analysis of the off-system activity here may provide converging evidence in support of the best fitting feedback mechanism arrived at through competitive model fitting (Chapter 4). The prediction here is that when the II system is the off-system there will be an increase in activity in the caudate. Similarly, when the RB system is the off-system there will be an increase in activity in the MTL.

5.3.2.2 Methods

The purpose of this analysis is to sort the trials in Experiment 1 and 2 based on the accuracy of the off-system and to use this sorting to examine corresponding functional activity in both the MTL and caudate ROIs. Toward this end, trials in Experiment 1 and 2 were defined based on whether the off-system was correct or incorrect regardless of the actual feedback delivered to the subject. The off-system is always defined as the system that was not assigned to the trial. For example, if an ‘A’ stimulus was evaluated by both the RB and II systems and the RB system judged it to be a ‘B’ and the II system judged it to be an ‘A’, the overt feedback to the subject would have been ‘incorrect’. If the RB system was assigned to this particular trial, the off-system (II here) would have been correct and the on-system (RB here) incorrect. All trials for all subjects in both experiments were processed similarly.

5.3.2.3 Results

In Experiment 1, there was no evidence of off-system activity either in the MTL or the caudate for II or RB subjects, respectively. In Experiment 2, there was evidence of off-system II activity only in the caudate ROI for RB subjects (Figure 5.13B). The accuracy and number of trials of the off-system responses is shown in Figure 5.13A. In general, the off-system was more accurate when it was associated with the appropriate subject group. In the RB group, when the RB system was the off-system, those trials were more accurate than when the II system was the off-system. The converse was true for the II group.

Given that we sorted the trials based on the II system prediction rather than the overt feedback given to the subject, there was a question of mis-match between internal and external feedback. Namely, how often did the correct feedback correspond to the II system's correct response and how often was the overt feedback inconsistent with the II system's prediction. The average number of overlapping trials for the RB subjects included in this analysis was 55% (SE = 0.024). That is, on average, the II system was receiving the same feedback as the RB system during only 55% of the trials. The other 45% of the trials were those in which the II system prediction was different from the RB system and the corresponding feedback differed from that given to the subject. The characterization of trials based on the off-system prediction differed in such a way as to illuminate functional activity within the caudate that was not observed in the previously described analysis of II system activity in RB subjects (see Figure 5.11).

5.3.2.4 Discussion

Pairing the predictions from a cognitive model such as PINNACLE with fMRI data here presents an opportunity to test assumptions about system interaction and feedback incorporation.

There are multiple forms of evidence implicating the caudate in II category learning, including the previously discussed fMRI data showing caudate activation in participants actively learning an II category structure (Nomura et al., 2007). Similarly, the MTL is more active when participants are acquiring an RB category structure. The MTL, then, can be considered to be a neural correlate of the on-system in RB subjects and the caudate the on-system in II subjects. This does raise the question, though, of what the off-system is doing when the on-system is actively acquiring a category representation.

The feedback structure of PINNACLE assumes that both the RB and II system generate predictions and incorporate feedback on each trial regardless of the system that was active on that trial. This means that the off-system is capable of developing a category representation while the on-system develops an independent representation. The capability of both systems to acquire the category representation may serve as a redundant mechanism in healthy humans, but when there is a challenge to one system, such as in a diseased state, the intact system could potentially operate alone.

The current analysis was designed to examine the incorporation of feedback by the off-system independent of the feedback to the on-system. The data from the RB subjects in Experiment 2 supports the idea that the II system, while not overtly responsible for behavior, is not only active in these subjects, but appears to be utilizing the feedback to update its category representation. The posterior regions of the body of the caudate that were more active for correct than incorrect off-system trials in these subjects are identical to the regions typically active during successful on-system II category learning. Thus, despite the success of the RB system in these subjects (and corresponding MTL activity), the II system appears to be operating in parallel through activity in the posterior caudate body.

While a similar result was not observed in the MTL when the RB system was the off-system in II subjects, this does not rule out the notion that both the RB and II systems are capable of learning simultaneously. Perhaps with a study designed specifically to test this idea of simultaneous learning systems, it would be possible to detect this off-system activity in the MTL. Regardless, the caudate activity observed in the off-system does serve as a form of convergent evidence that PINNACLE's mechanism of feedback returning to both systems may have some neural reality.

CHAPTER 6: IMPROVEMENTS TO PINNACLE AND FUTURE DIRECTIONS

6.1 Overview

In this section, the potential application of model-based fMRI analysis is discussed along with several mechanistic improvements to the RB and II systems within PINNACLE that would better reflect the neurobiological properties of the MTL and caudate. While these improvements are not currently part of PINNACLE, future work will depend on this type of conceptual framework.

6.2 Re-implementation of the RB system

One potential benefit of greater incorporation of brain imaging data in theories of category learning is that the additional constraints provided by neuroscience data may help unify the different cognitive modeling approaches. While there are many ways to computationally instantiate these models, there may be fewer ways to instantiate these models in a manner that captures the constraints from neuroimaging and neuropsychology. The involvement of the MTL and the caudate in particular will constrain theories of how the brain accomplishes category learning. The roles of these brain regions in memory research are very well-studied and they are among the most investigated brain regions at a neurobiological level.

The PINNACLE theory as currently implemented in a computational model has several demonstrated flaws. In future versions of the model, changes to the implementation of the PINNACLE theory should better equip the model to address hypotheses about these interactive brain systems. The major suggested modifications target the representational mechanisms by which the RB and II systems acquire and retain category information.

The neuroimaging results presented previously (fMRI Exp 1 and 2) support the idea that there are two learning processes that contribute to category learning that depend separately on the MTL and the caudate. While these results are interpreted in the context of the COVIS theory, that cognitive model does not make specific computational predictions about the representations inherent in the RB and II subsystems. With the suggested modifications to the PINNACLE model here, the cognitive processes described in the theory could be instantiated in such a way that would allow for more direct tests of the underlying hypotheses.

The RB system is currently set up to select a linear decision boundary, adjust it according to a particular learning rate, and with enough training, arrive at the optimal boundary (or matched to subject behavior). While this mechanism does serve to capture the behavior of the subjects in both Experiment 1 and 2, it may not be the best cognitive model of this system's operation. It seems particularly implausible that participants are actually manipulating a vertical line in their mind when learning these categories. Rather, a more realistic model is one in which exemplar stimuli close to the optimal category boundary are held in a memory buffer. A new stimulus is compared to this memorized exemplar and placed in either the same or the other category. The rule for categorization would then be something like: 'if stimulus X is less striped than the exemplar, it's an A, if it is more striped, it is a B'. The neuroimaging data suggests that this type of model may have neural reality based on the involvement of the MTL in the RB task. Declarative memory is known to depend materially on the structures within the MTL, and could be involved in the process of holding a particular stimulus in mind and/or recalling this stimulus for comparison purposes.

Implementing this modification to the RB system should be fairly straightforward. Rather than allowing the model to begin with a vertical line of some arbitrary frequency that is shifted according to negative feedback, a better model may be one that holds the frequency value of the most recent stimulus that is closest to category 'B' while still belonging to the 'A' category. As the model processes more stimuli, the exemplar will be discarded and replaced with a better example of the category. In effect, this boundary stimulus will be conceptually similar to the previous RB system, but will better reflect our notion of what the RB system is actually doing in this task.

6.3 Re-implementation of the II system

Currently, the II learning system represents the category structure by altering the intercept of a linear decision boundary. As in the RB system, it is similarly implausible here that the subject is internally manipulating a linear boundary. A better mechanism for the II learning system may be a clustering learning algorithm similar to the Striatal Pattern Classifier (SPC) model (Ashby and Waldron, 1999). This is an example of a model in which regions of perceptual space are labeled by II category learning, a process hypothesized to involve cortico-striatal circuits through the posterior regions of the caudate.

To implement something like the SPC in the II system of the current PINNACLE model would require a number of structural changes. First and most importantly, the linear decision bound would no longer be used to track the internal representation of category structure. Instead, striatal units in stimulus space learn to associate particular stimuli with an output response given the model's experience. Categorization behavior should emerge from this structure as the model learns in the form of a 3-dimensional topographical 'map' of the space (an example is shown in Figure 6.1). In Figure 6.1A, the stimulus space shows an example of a representation that is early in learning. Each cluster is fairly small and the boundary between the categories is not clear. Later in learning, these clusters grow in size (Figure 6.1B) presumably due to the positive feedback associated with learning. From this arrangement of clusters, the category boundary becomes more apparent.

Instantiating this type of II system would involve implementing an ‘attractor network’ that has previously been shown to be successful in models of language learning (Maddox et al., 2002) and vision research (Marr and Poggio, 1976). The specifics of the attractor networks are described elsewhere (Zemel and Mozer, 2001). The basic idea of an attractor network is that an input pattern is drawn toward the node (or striatal units in the SPC model) in stimulus space that is closest in proximity. The shape of the stimulus space is determined by experience, so the model does not need any prior knowledge of the structure. With each incoming stimulus and accompanying feedback information, its similarity to previously experienced stimuli will either be drawn toward an existing node or will create a new node. Over many trials, the shape of the category space should reflect the learning that has occurred in a manner similar to the perceptual shaping curve in Figure 4.2. Areas of stimulus space that have been repeatedly endorsed as belonging to the same category will grow and encapsulate other areas that require the same motor response. In this manner, II category learning will be represented as something closer to perceptual learning rather than a linear decision boundary.

6.4 Potential extrapolations of PINNACLE

Neuroimaging of RB and II category learning successfully identified two dissociable neural networks supporting these cognitive activities. We can hypothesize about the mechanism(s) by which this learning takes place in these systems, but it is difficult to examine them with fMRI alone. Computational and mathematical modeling, in contrast, are constrained by specific hypotheses about the underlying mechanisms supporting cognitive operations. Applying the known biological constraints of the neural systems in question to models of cognition has the potential to advance our understanding beyond the level of dissociating neural activity with fMRI.

The potential changes to the RB and II systems outlined here better reflect our current understanding about the underlying cognitive and neural mechanisms guiding this behavior. The fact that the changes differ substantially between the systems is also a deliberate choice that supports the notion that these systems represent very different learning processes. Further, the development and improvements to PINNACLE opens up the possibility of extending the framework to other types of paradigms both within categorization and to other kinds of learning. Because the model is based on the basic notion of modeling 2 learning systems, it should be flexible enough to learn in other types of skill learning paradigms that may rely on similar neural machinery, for example, the SRT task, weather prediction task, or route navigation tasks.

The SRT task has been shown to recruit different cortical and sub-cortical regions during motor skill learning depending on whether the learning is accompanied by awareness (Willingham et al., 2002). Neuroimaging of route navigation suggests the involvement of both the MTL and caudate depending on the type of learning strategy employed (Iaria et al., 2003). Similarly, neuroimaging of the weather prediction task has demonstrated a reliance on the operation of both the MTL and the caudate of the basal ganglia depending on the cognitive strategy employed by the subject. When relying on the probabilistic feedback, there was increased involvement of the caudate, whereas an explicit memorization strategy elicited MTL activity (Poldrack et al., 2001). Based on the similarity in neural systems engaged across these different tasks, it seems plausible that PINNACLE would eventually be able to make predictions of trial-by-trial system participation based on these data.

The weather prediction task may be the most feasible option for application to PINNACLE. In this task, a series of cues are probabilistically associated with one of two outcomes. Based on the cues that appear on each trial, the participant must guess which outcome is predicted. As discussed previously, amnesic patients are able to learn these associations normally (presumably relying upon their intact basal ganglia system) while patients with damage to the striatum are impaired. Patients with a milder form of Parkinson's disease, however, have shown the ability to learn this task almost as well as healthy control subjects. Further, functional imaging of patients with mild Parkinson's disease revealed activity in the MTL, in contrast to control subjects who showed activity in the striatum (Moody et al., 2004). Given the apparent interchangeability of the MTL- and basal ganglia-driven learning during this task, the current implementation of PINNACLE is well suited to model this task. A prediction of this potential application is that early in learning the weather prediction task the RB system would dominate learning. That is, participants would explicitly attempt to memorize specific examples or arrive at a rule that maximized their accuracy in the task. Over extended training, however, the operation of the II system would eventually come online as it incorporated the feedback information. The overall behavioral profile of a participant in the weather prediction task would thus look similar to that of a person learning an II category structure. PINNACLE could also be used to fit data obtained from neuropsychological populations. For example, should PINNACLE find that the MTL-based RB system dominated trials in mild PD patients, this would provide convergent evidence that the intact declarative memory system was driving their learning in this task.

Should this model prove flexible enough to learn other types of categorization or skill learning paradigms, it would be interesting to extrapolate to other types of cognitive phenomena above and beyond learning. For example, can the model account for the development of expertise? Some suggest that during learning, a particular network is in operation, but once a task has been learned up to the point of exhibiting expert behavior, this network drops out and another ‘expertise’ network comes online. By overtraining participants in the sine-wave grating paradigm here, and scanning both at the beginning of training and after the categories are extremely well learned, one could test this hypothesis. If true, the RB system would be highly active as the participant works to figure out the rule that defines the categories. As the category division becomes increasingly obvious, the responses should rely less on feedback. At some level of high performance, the RB system is no longer necessary and a more automatic, efficient processing loop is responsible for guiding responses. This function may be subserved by the posterior II circuitry or perhaps through direct cortico-cortico connections from PFC to motor cortex. The SPEED model of automaticity (Ashby et al., 2007) is one recent example of such a model, but unlike PINNACLE, this model does not attempt to account for the multi-system competition that may occur early in learning.

6.5 Conclusions

Observed dissociations amongst patient groups challenged with different learning tasks initially established the neuroanatomical independence of the MTL- and basal ganglia-based memory systems (Chapter 1). Behavioral data and evidence from animal studies bolstered this claim and more recently, neuroimaging has provided a method for visualizing the functional operation of these networks in healthy people. Studying the brain activity of healthy individuals is complicated by the fact that it is difficult to control the way in which people learn. That is, people presented with a learning task, even when given explicit instructions, are capable of using all the cognitive resources they have available. Under the multiple memory systems hypothesis, different types of memory and the corresponding neural networks can be engaged on any learning task. The challenge then is to dissociate these learning systems in the brain.

In the two neuroimaging experiments described in Chapter 2, we demonstrated a dissociation between successful RB and II category learning activity in the MTL and caudate respectively. The involvement of these neural structures is consistent with our understanding of the cognitive function of each of these regions and with the predictions of the COVIS theory of category learning. The data here thus provide one of the first demonstrations of the utility of coupling fMRI with cognitive models of categorization.

Neuroimaging subtractions typically rely on the classification of behavior, usually via accuracy or some objective measure of performance. As such, an inherent assumption in the accuracy-based analysis of these data is that learning in the RB subjects depended upon the consistent application of an RB strategy and learning in the II subjects on an II strategy. Yet as we demonstrated with both the mathematical DBT-models (Chapter 3) and PINNACLE trial-by-trial predictions (Chapter 4), both subject groups exhibited a mixture of strategies throughout learning.

Using model-based predictions to organize neuroimaging data can thus improve upon the existing analysis techniques by isolating activity associated with strategy-specific behavior. Rather than grouping activity based on the imposed category structure, a mathematical characterization of strategy-use can expose subjects who are using an inappropriate strategy or assess how well the appropriate strategy is being utilized. Using this block-sorting technique, in the mathematical model-driven organization of fMRI data, we identified a number of regions not observed in the accuracy-based fMRI analysis. Specifically, PFC and visual association cortex activity was associated with blocks that were best fit with RB and II strategies, respectively. These regions are thought to function in conjunction with the operation of the MTL and posterior body of the caudate in RB and II learning as described initially in COVIS. The visualization of these additional regions would not have been possible without the application of these models to the fMRI data.

The shift from static DBT-style RB and II models to a more sophisticated model of category learning that allows both systems to interact was motivated by the hope that the sensitivity of trial-by-trial strategy predictions would improve the specificity of the resulting fMRI contrasts. The framework of PINNACLE (Chapter 4) depends on the notion that there are 2 systems active in the normal healthy brain; an RB system dependent upon the MTL and an II system dependent upon the reciprocal loops through posterior regions of the caudate and visual association cortex. Mediation between these systems is assumed to rely on a decision module situated in the PFC. Neuroimaging analyses aimed at identifying activation associated with this component of PINNACLE has provided convergent evidence that the PFC is involved in this system mediation (Chapter 5).

From the group and individual subject modeling analyses, we've demonstrated both the validity of PINNACLE as a model of 2 interactive learning systems in the brain as well as its ability to account for the data in 2 different categorization studies. In this stage of assessing the model's behavior, a number of assumptions were called into question. Through competitive model fitting of different versions of PINNACLE, we determined that the RB and II systems require distinct learning rates and that the data is best accounted for with a feedback mechanism that provides both systems with positive and negative feedback (Chapter 4). This feedback structure is in line with the notion that the RB and II systems operate simultaneously, yet does not rule out the possibility that one system can inhibit the other or one system can enhance the confidence of the other. Although, some evidence for the reality of this feedback mechanism comes from an examination of off-system activity in RB subjects. Despite the successful utilization of an RB strategy in these subjects, the posterior regions of the caudate were active in conjunction with PINNACLE-identified II off-system success (Chapter 5).

Together, the collection of neuroimaging and modeling data discussed here demonstrate the utility of combining these two methodologies. The symbiotic relationship between model development and fMRI data analysis has advanced our understanding of RB and II category learning systems in the brain.

CHAPTER 1 FIGURES

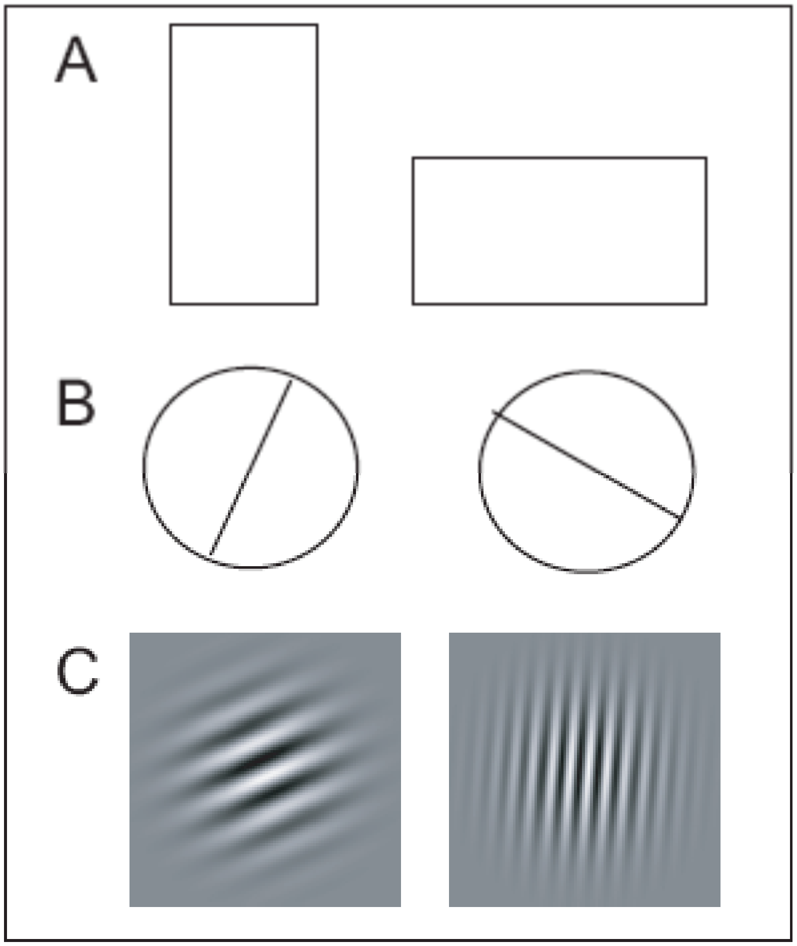


Figure 1.1: examples of stimuli in category learning paradigms

(A) Rectangular stimuli that vary in width and length (B) Circular stimuli that vary in diameter and line orientation (C) Sine wave stimuli that vary in frequency and orientation.

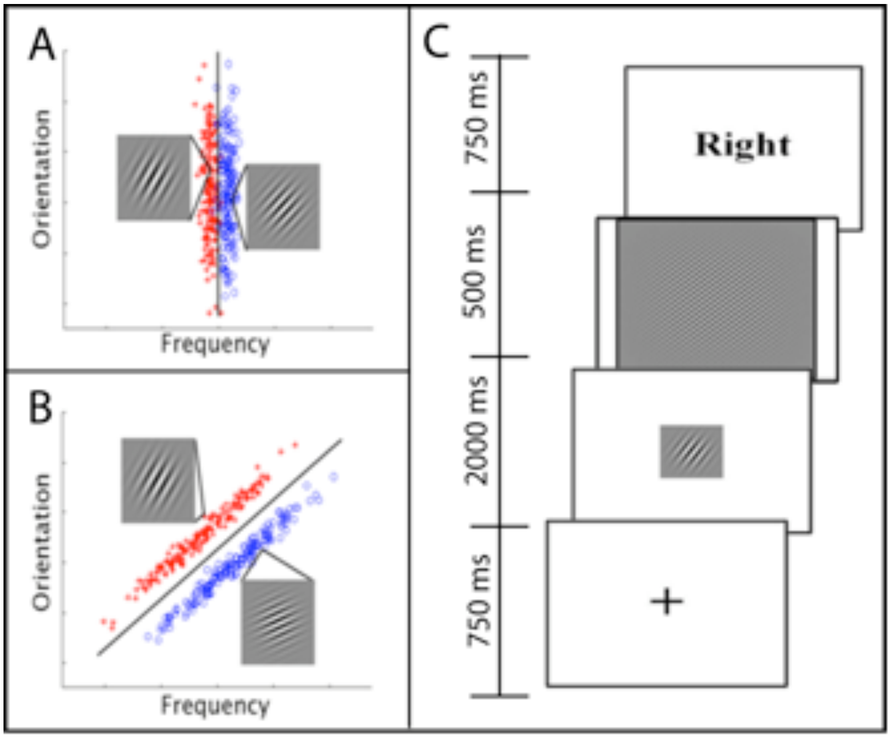


Figure 1.2: Exp.1 sine wave stimuli

RB (A) and II stimuli (B). Each point represents a distinct Gabor patch (sine-wave) stimulus defined by orientation (tilt) and frequency (thickness of lines). In both stimulus sets, there are 2 categories (red and blue points). RB categories are defined by a vertical boundary (only frequency is relevant for categorization) whereas II categories are defined by a diagonal boundary (both orientation and frequency are relevant). In both RB and II stimuli there are examples of a stimulus from each category. (C) Schematic of a single trial. A fixation point is followed by the to-be-categorized-stimulus (either RB or II depending on the subject), then a short visual mask that is followed by the feedback. The subject responded ‘category A’ or ‘category B’ during the 2 seconds the stimulus was on the screen using hand-held buttons. The length of the inter-trial interval (ITI) was pseudorandom and based on between zero and five 4-s “fixation-only” trial periods arranged to maximize the separability of the measured hemodynamic response to stimulus trials.

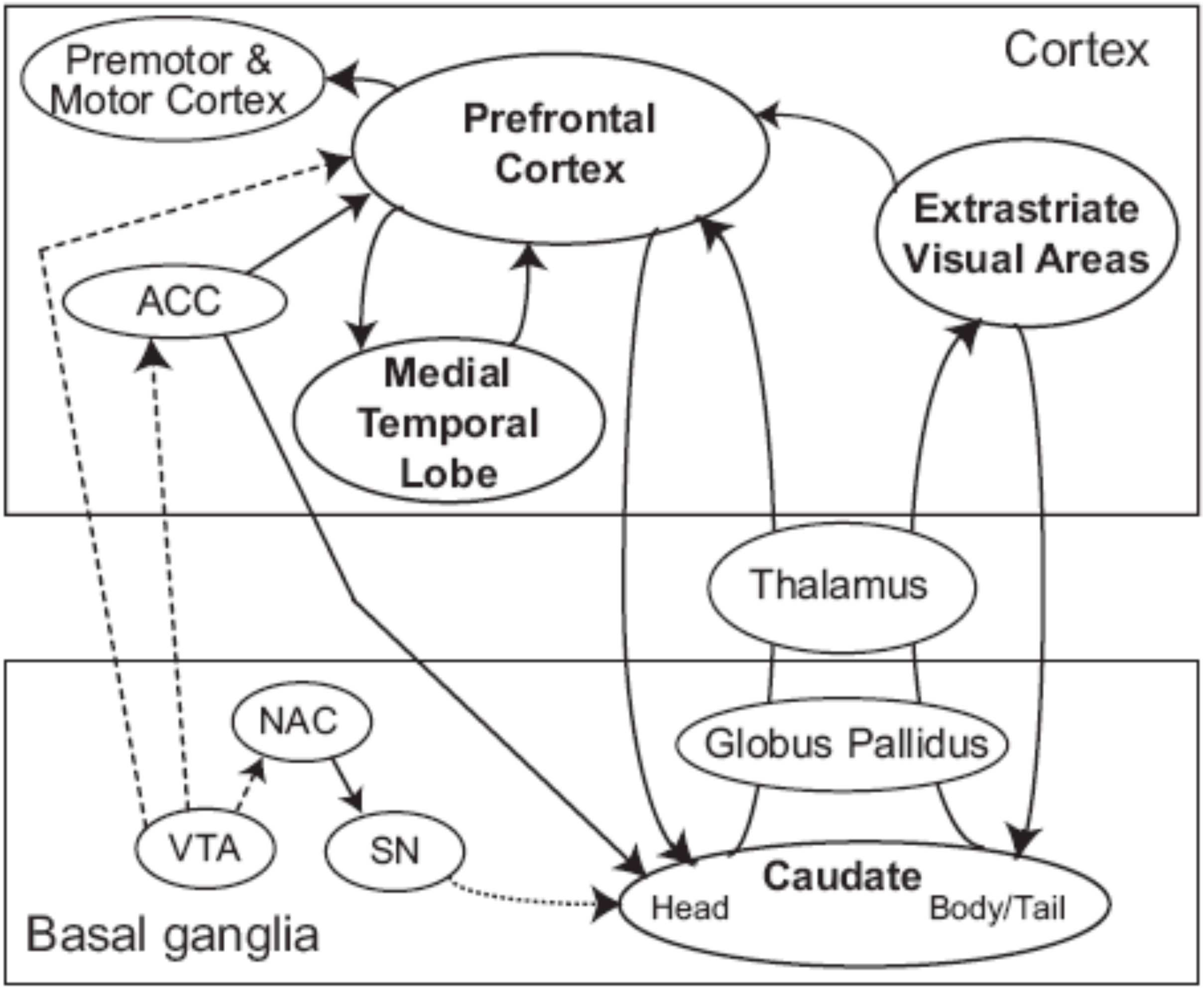


Figure 1.3: Adapted COVIS schematic

Schematic of COVIS model adapted from Ashby, Alfonso-Reese et al. (Ashby et al., 1998) to incorporate reciprocal connections between the PFC and MTL (reflecting interactions found in research on declarative memory) as well as between extrastriate visual cortex and posterior regions of the caudate. RB category learning is hypothesized to depend on connections between the head of the caudate, PFC, the MTL and the ACC. II category learning is hypothesized to depend on changes within extrastriate cortex supported by the body/tail (posterior) regions of the caudate. Dotted lines represent dopaminergic projections for incorporating feedback in learning. ACC (anterior cingulate cortex); NAC (nucleus accumbens); SN (substantia nigra); VTA (ventral tegmental area).

CHAPTER 2 FIGURES

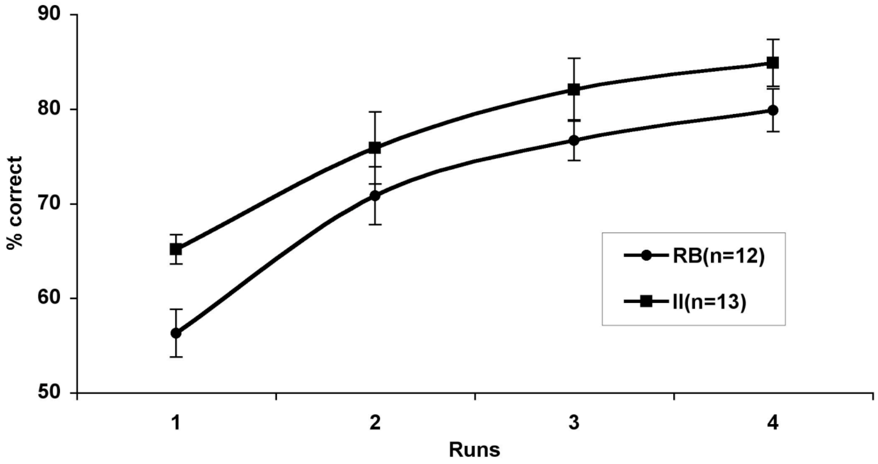


Figure 2.1: Exp. 1 learning curves

Experiment 1. Accuracy of RB and II “learners” across 4 runs. Both groups performed above chance in all runs and average accuracy did not differ between groups (F(1,24) = 1.25, n.s.). RB mean performance across runs was 70.8% (SE =1.83) and II was 74.9% (SE = 3.03).

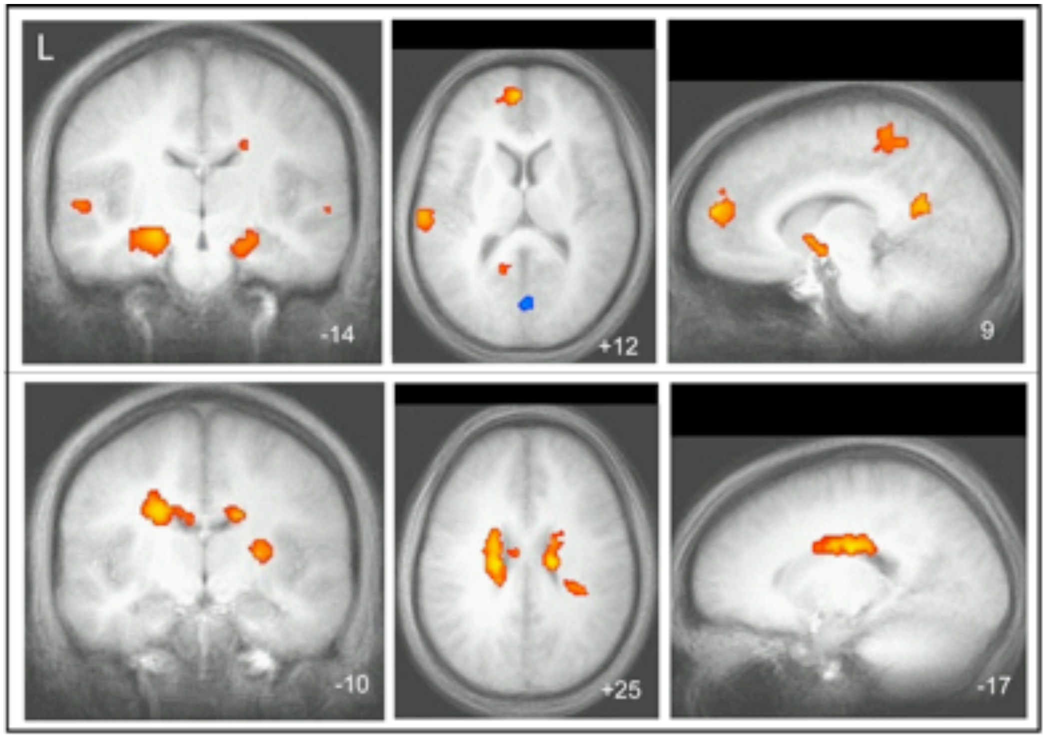


Figure 2.2: Exp. 1 whole-brain fMRI data

Experiment 1. (top) A portion of the successful RB categorization network featuring the MTL ($t(11) > 4.5$, cluster > 300 mm^3). Sagittal views are of the right hemisphere. (bottom) A portion of the successful II categorization network featuring the body of the caudate. ($t(12) > 4.5$, cluster > 300 mm^3).

Table 2.1: Exp. 1 RB and II correct/incorrect differences

Experiment 1: Volume-thresholded clusters of areas that activated differentially to correct versus incorrect categorization trials for the RB and II groups. T>4.5, cluster>300mm^3

Brain Region	Brodmann's Area	Talairach Coordinates (x, y, z)	Cluster size (mm^3)
RB group:			
Correct > Incorrect			
Left anterior hippocampus		-18, -8, -12	2172
Right anterior hippocampus		16, -5, -10	1781
Left superior frontal gyrus	8	-20, 17, 53	1688
Left medial frontal gyrus	10	-8, 53, 19	1203
Left transverse temporal gyrus	41	-59, -21, 10	1109
Right caudate body		18, -26, 28	1094
Left paracentral lobule	5	-8, -41, 56	1078
Right superior temporal gyrus	22	59, -8, -2	578
Left posterior cingulate	30	-11, -58, 17	531
Right precuneus	7	13, -47, 60	453
Right paracentral lobule	6	4, -34, 58	406
Right superior frontal gyrus	6	2, 2, 62	406
Incorrect > Correct			
Right cuneus	18	2, -78, 5	719
Right supramargninal gyrus	40	55, -55, 37	578
II group:			
Correct > Incorrect			
Left caudate body		-20, -14, 29	2766
Left lentiform nucleus		24, -7, 3	1656
Right caudate body		17, -11, 28	969

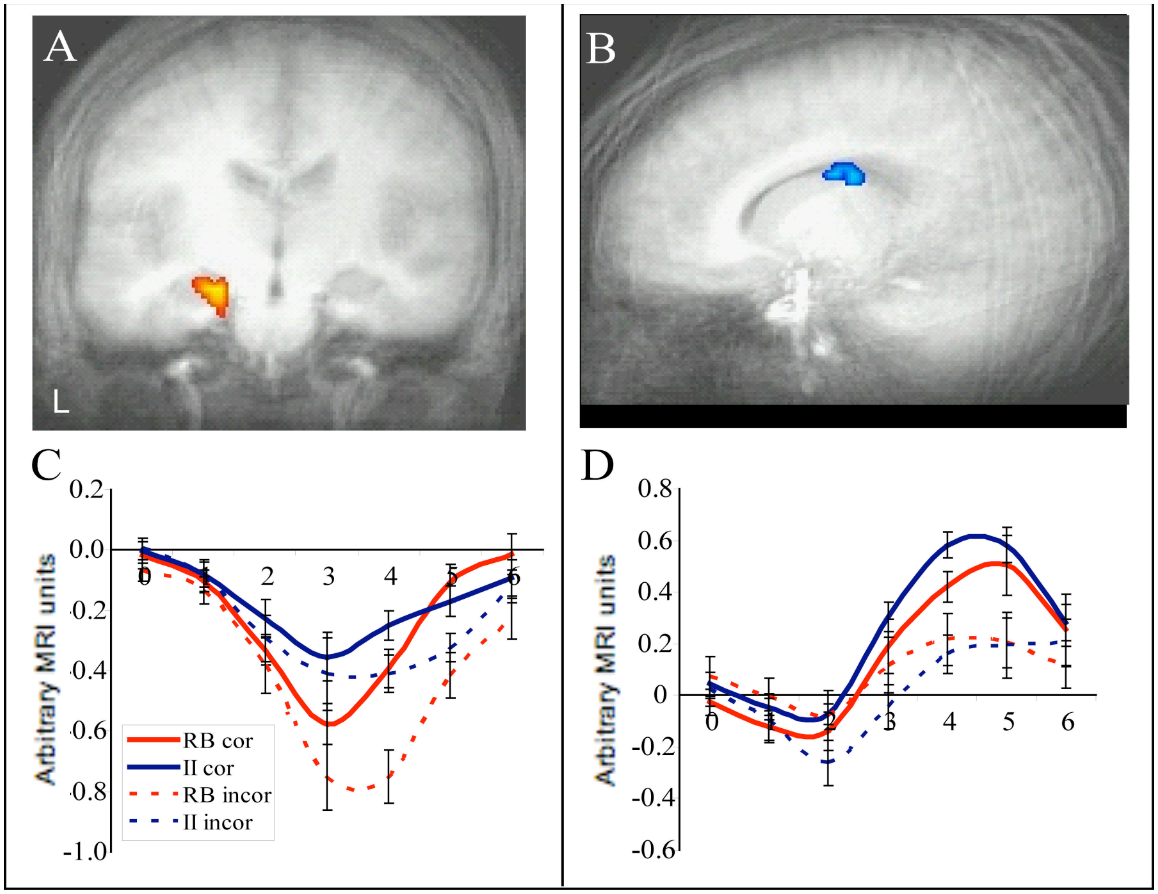


Figure 2.3: Exp. 1 ROI fMRI data

Experiment 1. (A) RB categorization > II categorization in the left MTL (t (24) > 2.0, cluster > 700 mm^3). (B) II categorization > RB categorization in the right body of the caudate (t (24) > 2.0, cluster > 600 mm^3). ROI-AL methodology (Stark and Okada, 2003) was used to align regions of interest in both A and B. Note that cross-subject ROI alignment was improved at the cost of whole-brain alignment. (C) Peri-stimulus time (PST) course for the left MTL ROI. (D) PST course for the right caudate ROI.

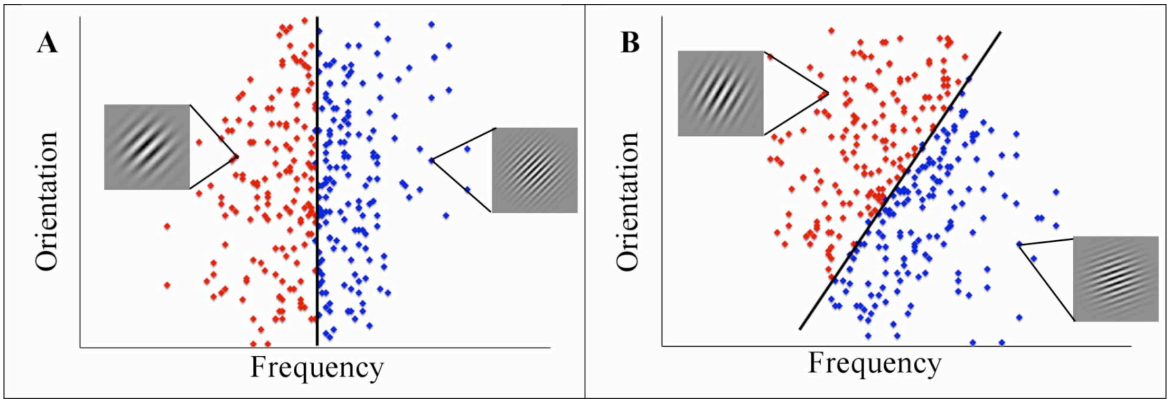


Figure 2.4: Exp. 2 stimuli

Experiment 2. (A) RB and (B) II stimuli. Each point represents a distinct Gabor patch (sine-wave) stimulus defined by orientation and frequency (thickness of lines). In both stimulus sets, there are 2 categories (red and blue points). RB categories are defined by a vertical boundary (only frequency is relevant for categorization) whereas II categories are defined by a diagonal boundary (both dimensions are relevant). In both RB and II stimuli there are examples of a stimulus from each category.

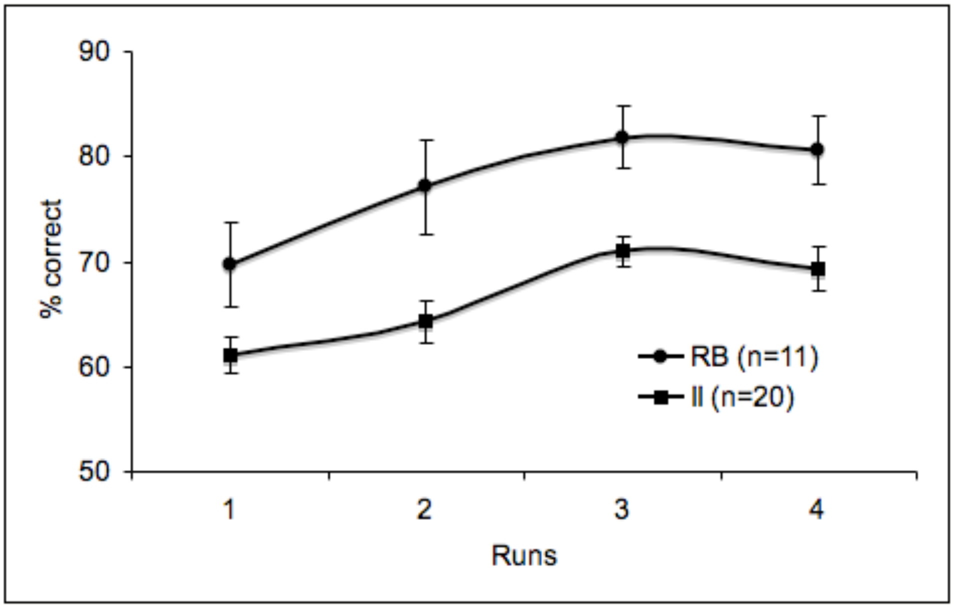


Figure 2.5: Exp. 2 learning curves

Experiment 2. Accuracy of RB and II “learners” across 4 runs. Learning across runs was reflected in a significant linear trend (F (1, 29) = 24.716, $p < 0.05$). Mean accuracy averaged across all 4 runs for the RB group was 77.3% (SE = 0.034) and for the II group was 66.6% (SE = 0.055). The RB group accuracy was significantly greater than the II accuracy across all 4 runs (F (1,29) = 12.5, $p< 0.05$).

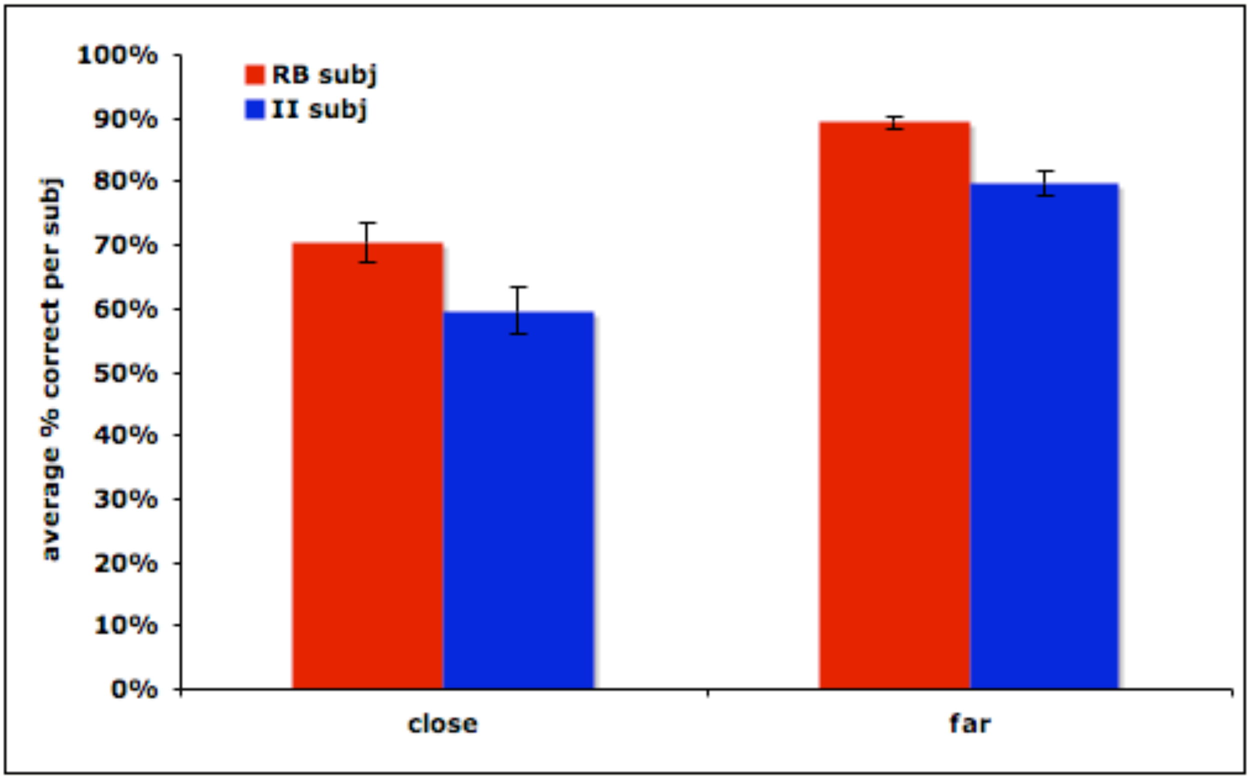


Figure 2.6: Exp. 2 accuracy vs. distance

Accuracy of RB and II "learners" across 4 runs as a function of distance of the stimuli from the bound. The same pattern is observed in both RB and II subjects that far stimuli are more accurate than close stimuli.

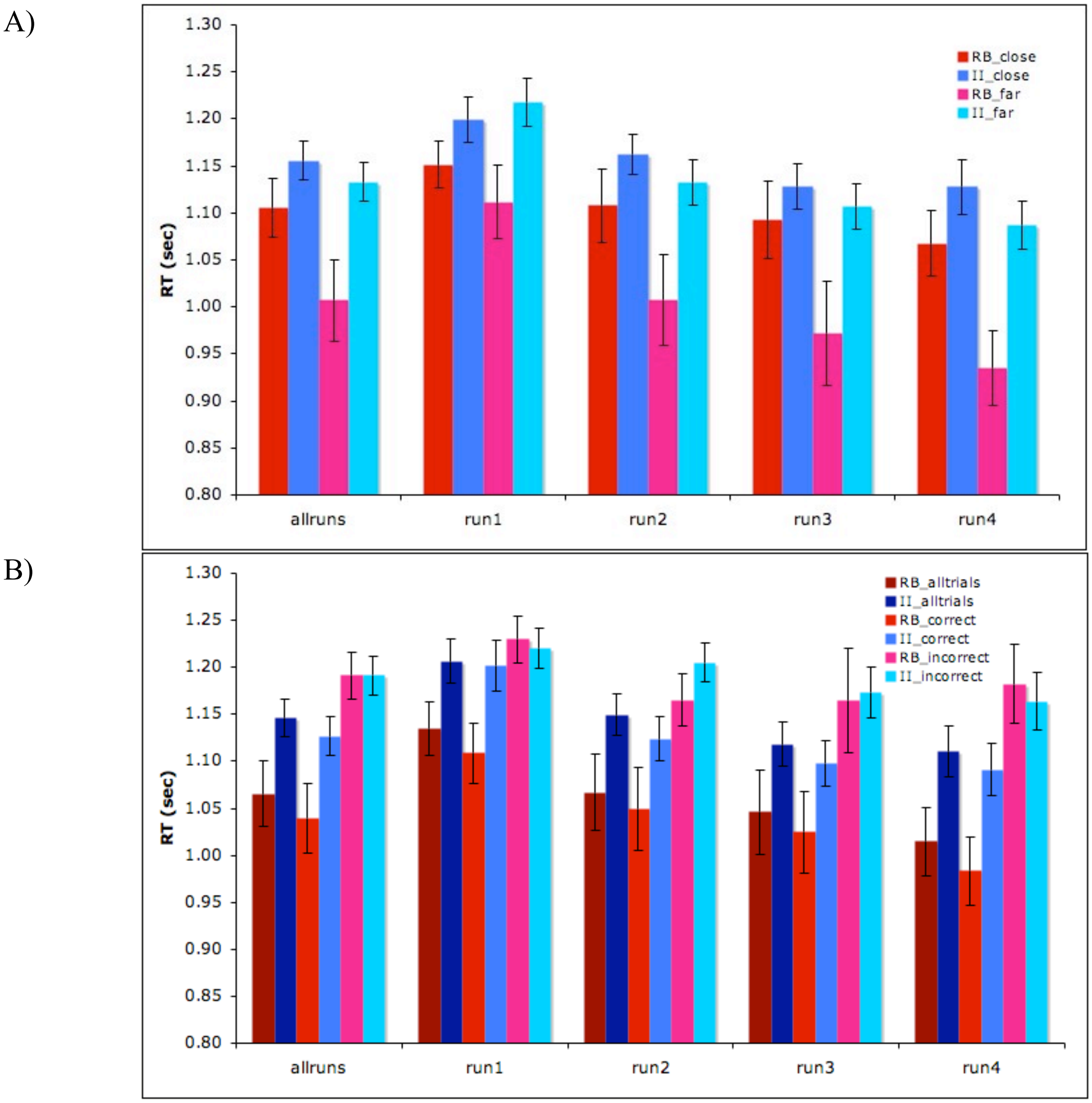


Figure 2.7: Exp. 2 RT differences within trial types and between groups

(A) RT differences between RB and II groups for correct and incorrect trials. (B) RT differences between RB and II groups for close and far trials.

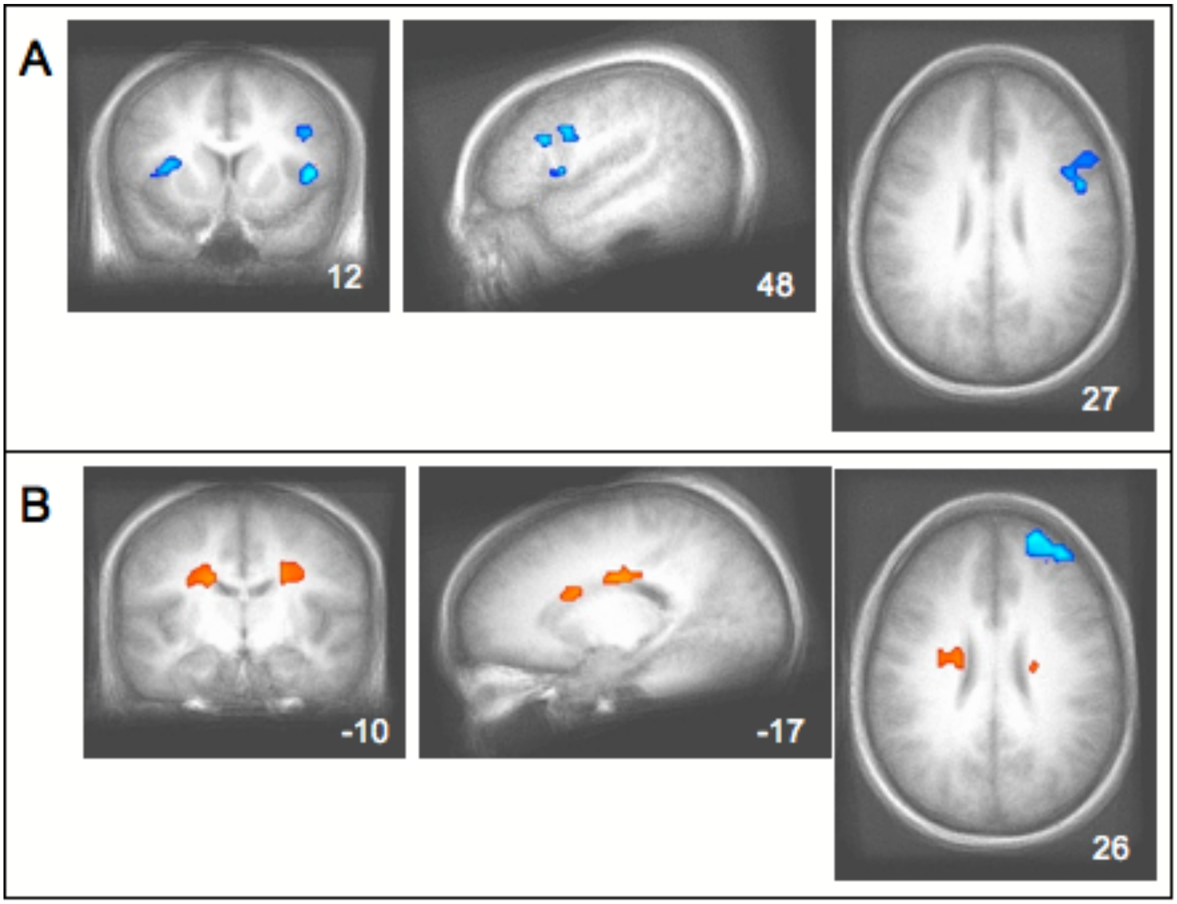


Figure 2.8: Exp. 2 whole-brain fMRI data, RB and II group subtraction

(A) RB and (B) II group correct vs. incorrect activity differences across all 4 runs in Experiment 2.

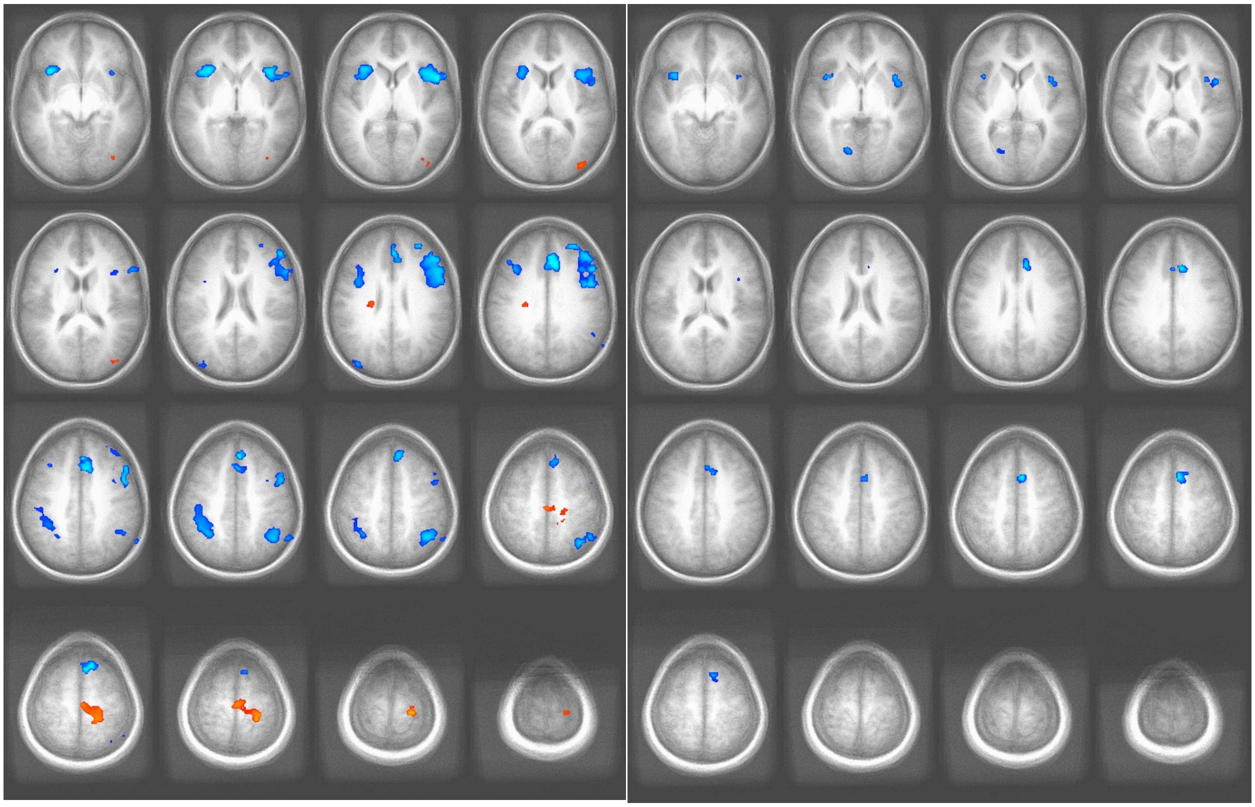


Figure 2.9: Exp. 2 close and far trial activity in the RB subject group

Correct vs. incorrect trials in RB subjects for (left) close and (right) far trials in Experiment 2.

Table 2.2: Exp. 2 RB correct/incorrect differences within close and far trial types

Experiment 2: Volume thresholded clusters of areas that activated differentially to correct versus incorrect close and far categorization trials for the RB and II groups (all runs). T>4, cluster>350mm^3

Brain Region	Brodmann's Area	Talairach Coordinates (x, y, z)	Cluster size (mm^3)
RB group (close trials):			
Correct > Incorrect			
Right precentral gyrus	44	14, 29, 60	3578
Right middle occipital gyrus	19	36, -82, 10	875
Incorrect > Correct			
Right middle frontal gyrus	9	46, 19, 30	8297
Right precentral gyrus	44	42, 15, 7	4141
Left inferior parietal lobule	40	-37, -42, 42	3547
Right medial frontal gyrus	6	5, 29, 36	3250
Left insula	13	-32, 20, 4	3250
Right inferior parietal lobule	40	40, -56, 46	3016
Left middle frontal gyrus	9	-36, 16, 30	1391
Right superior frontal gyrus	6	9, 21, 56	859
Right superior frontal gyrus	9	28, 45, 30	500
Right inferior parietal lobule	40	59, -58, 41	484
Left superior occipital gyrus	19	-38, -85, 25	391
RB group (far trials)::			
Incorrect > Correct			
Right insula	13	40, 9, 6	1125
Right superior frontal gyrus	6	7, 12, 50	1031
Right cingulate	32	9, 23, 32	750
Left insula	13	-35, 14, -1	594
Left lingual gyrus	18	-16, -68, 3	406

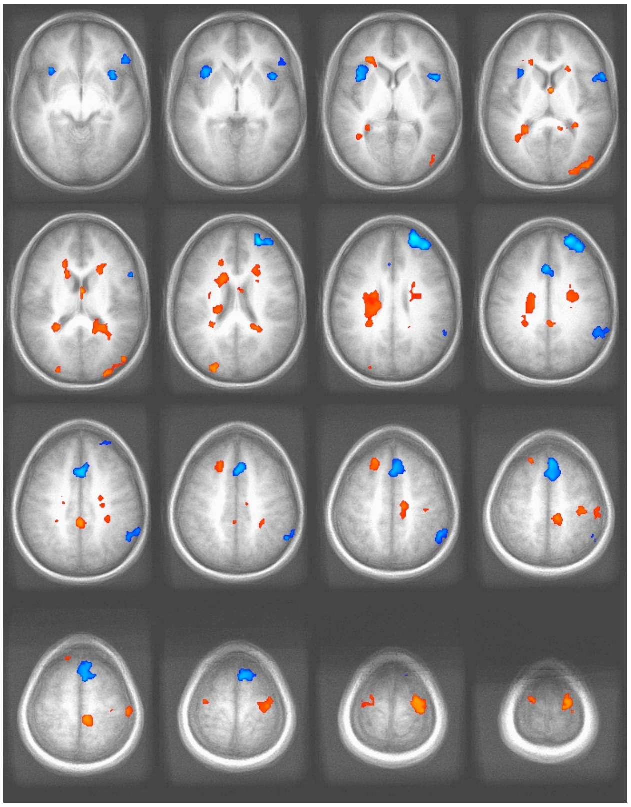


Figure 2.10: Exp. 2 Close correct vs. incorrect trials in II subjects

Correct vs. incorrect trials in II subjects for close trials. No successful categorization activation differences were observed in II subjects for the far trials.

Table 2.3: Exp. 2 II correct/incorrect differences within close and far trial types

Experiment 2: Volume thresholded clusters of areas that activated differentially to correct versus incorrect close and far categorization trials for the RB and II groups (all runs). T>4, cluster>350mm^3

Brain Region	Brodmann's Area	Talairach Coordinates (x, y, z)	Cluster size (mm^3)
Correct > Incorrect			
Left insula	13	-25, -28, 24	6406
Right superior frontal gyrus	8	4, 19, 49	6078
Right precentral gyrus	6	26, -18, 68	2359
Left caudate body		-18, 20, 17	2031
Right cingulate gyrus		25, -8, 31	1828
Right posterior caudate		22, -44, 18	1750
Right middle occipital gyrus	18	38, -85, 13	1609
Right paracentral gyrus		10, -32, 53	1609
Left superior frontal gyrus	8	-19, 27, 47	1203
Right anterior cingulate gyrus		22, 21, 18	1062
Left cingulate gyrus	31	-1, -38, 36	625
Left precentral gyrus	6	-25, -15, 68	625
Right postcentral gyrus	2	53, -26, 55	562
Right inferior parietal lobule		29, 36, 39	562
Left middle occipital gyrus	19	-25, -88, 21	547
Right precentral gyrus		36, -25, 57	484
Right caudate		2, -3, 16	391
Incorrect > Correct			
Right superior frontal gyrus	9	27, 50, 28	3438
Right inferior parietal lobule	40	56, -51, 39	2297
Left insula	13	-34, 18, 5	1766
Right inferior frontal gyrus		50, 31, -2	547

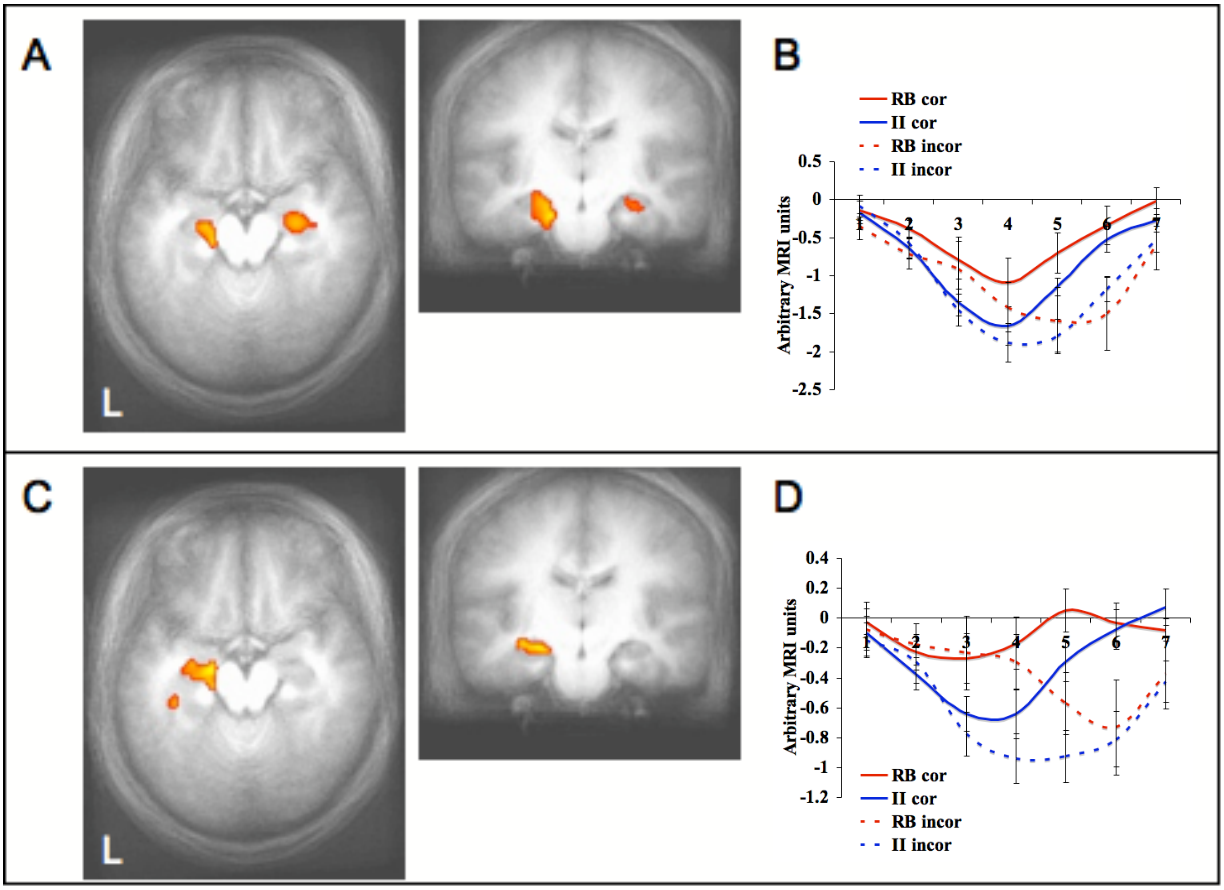


Figure 2.11: Exp. 2 MTL ROI fMRI data

Correct vs. incorrect trial activity in the MTL ROI (A) in RB subjects and (C) II subjects. The peri-stimulus time (PST) course for the left MTL cluster in (B) RB and (C) II subjects.

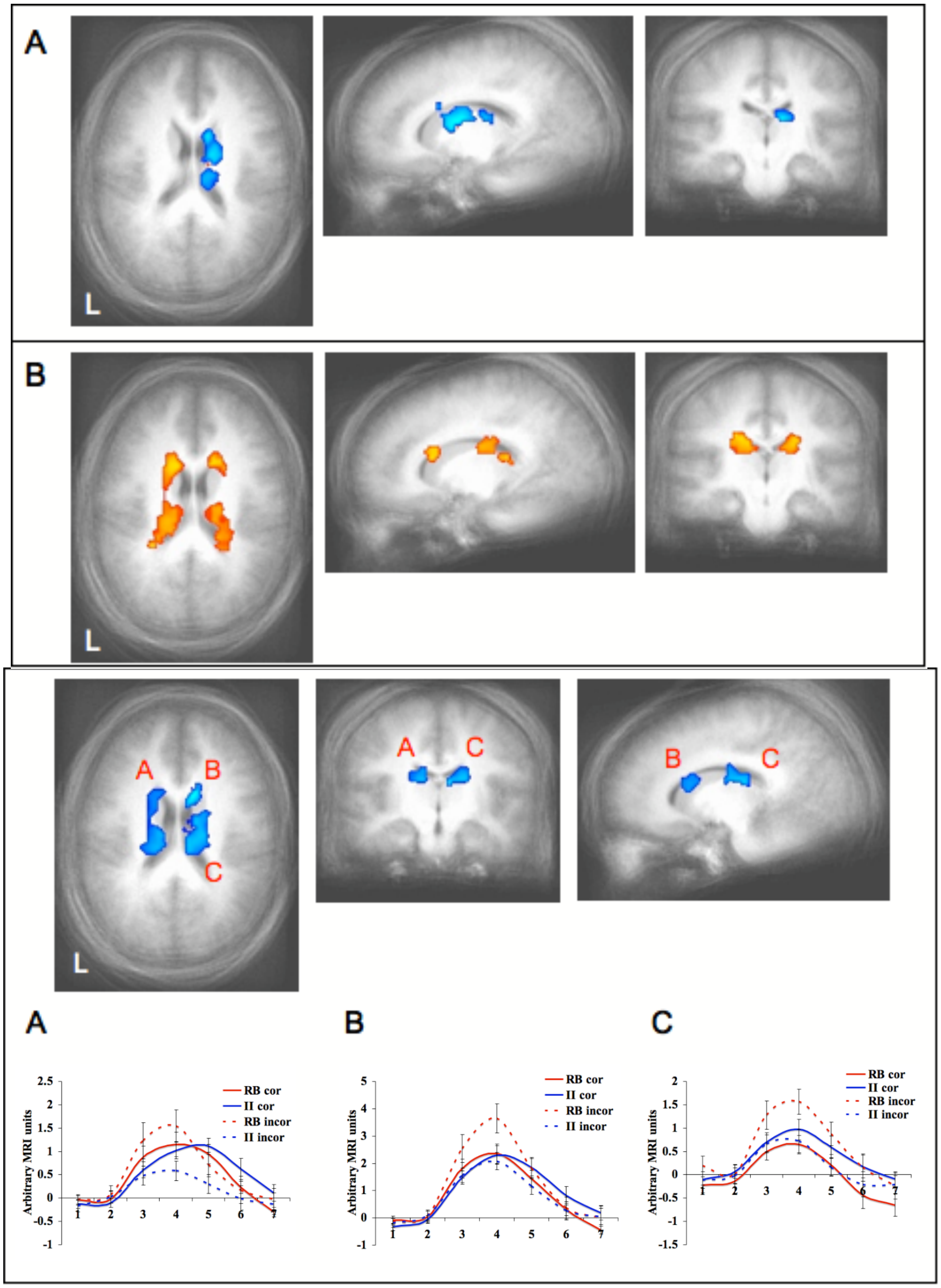


Figure 2.12: Exp. 2 Caudate ROI fMRI data

Correct vs. incorrect trial activity in the caudate ROI (A) in RB subjects and (B) II subjects. The double subtraction of RB and II subject groups is shown at the bottom. The corresponding peri-stimulus time course for three significant clusters of activity are shown below the double subtraction images.

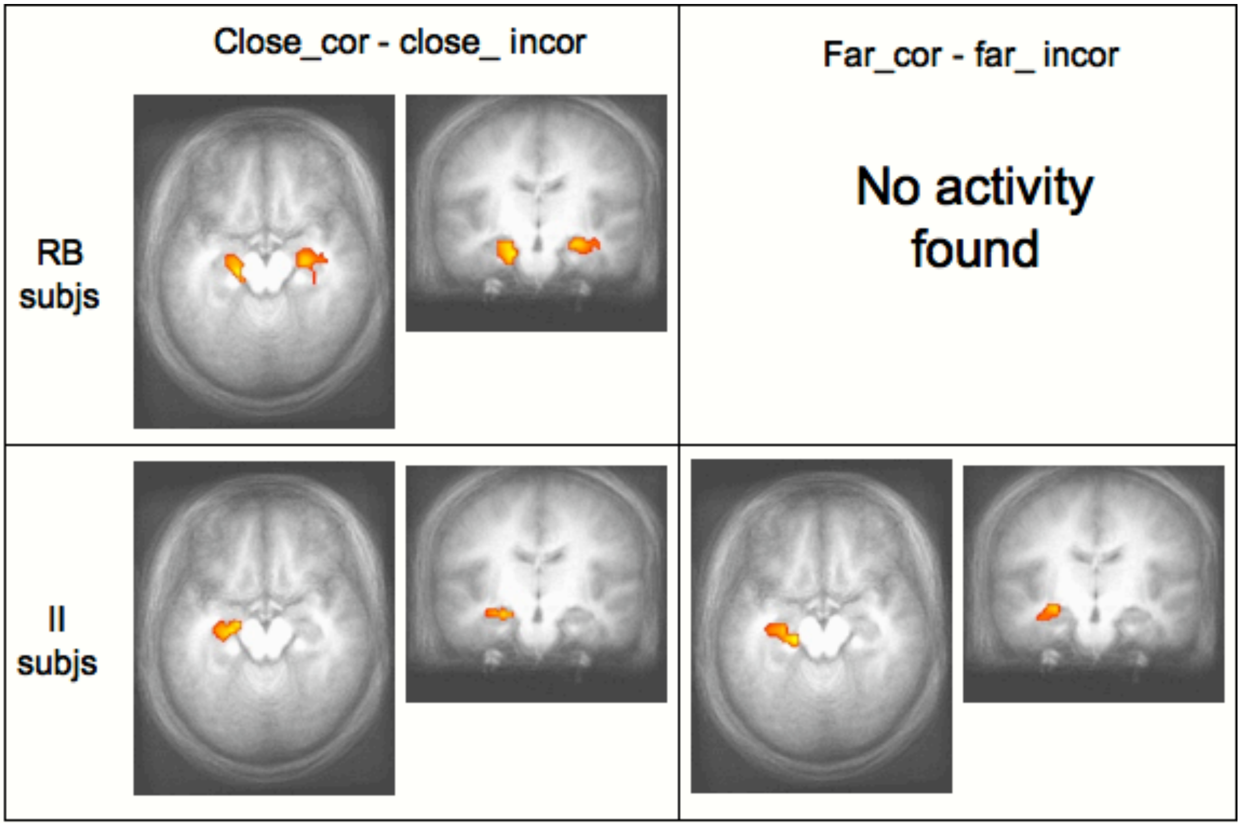


Figure 2.13: Exp. 2 MTL ROI

Correct vs. incorrect trials in the MTL ROI for close and far trials. There does not appear to be a difference in activity between the close and far trials in the II subject group. The far trials in RB subjects do not show any successful categorization difference, but incorrect far trials are relatively scarce in this subtraction.

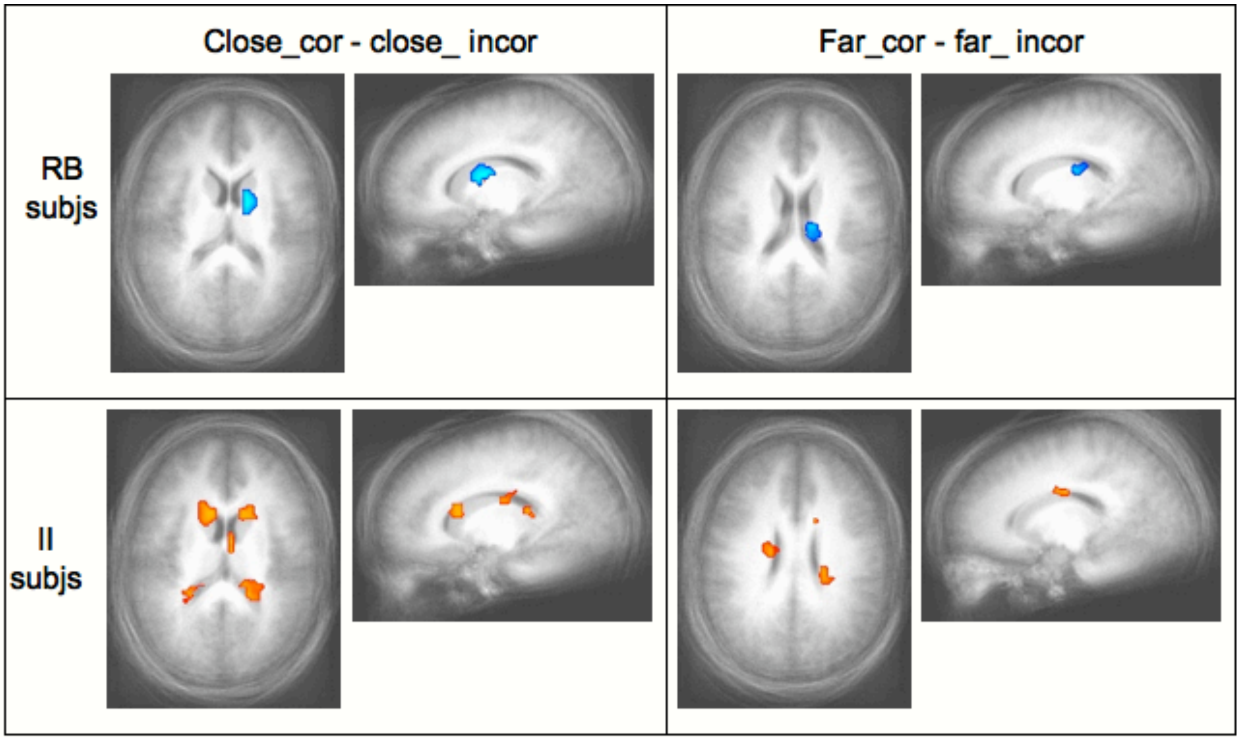


Figure 2.14: Exp. 2 Caudate ROI

Correct vs. incorrect trials in the caudate ROI for close and far trials. The difference between correct and incorrect trials is similar for the II group both in close and far trial types. In the RB group, far trials seem to elicit differential activity in the posterior region of the caudate whereas close trials are associated with a cluster in anterior caudate body. This difference is greater for incorrect than correct trials.

CHAPTER 3 FIGURES

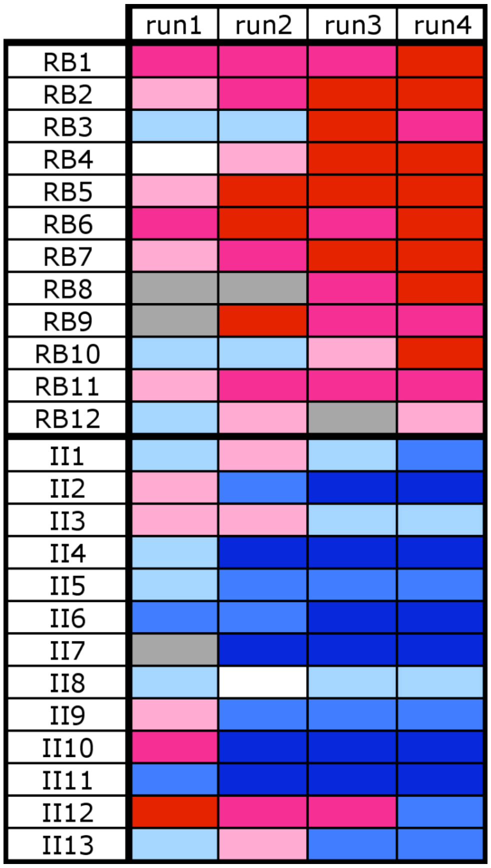


Figure 3.1: Exp. 1 Mathematical model fits for RB and II subjects

Experiment 1 model fits for RB (top) and II subjects (bottom). Fit values were separated into thirds with increasing hue indicating better fit. RB fits are in warm and II fits in cool colors. The darkest hues were included in the fMRI analysis in Figure 3.2.

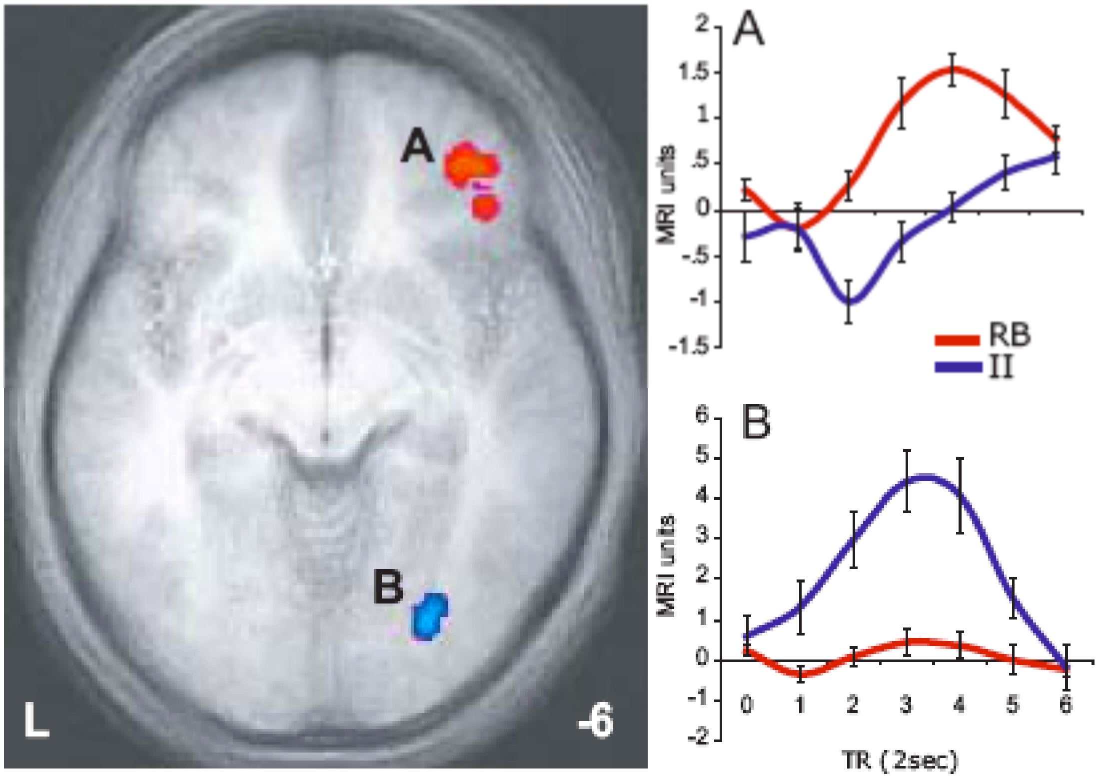


Figure 3.2: Exp. 1 model-identified best RB and II fMRI activity

Experiment 1. Correct trial activity for model-identified best strategy use and the corresponding peri-stimulus time courses of activity within significant clusters. (A) Best RB fit runs evoked greater activity than best II fit runs in right PFC. (B) Best II fit runs evoked greater activity than best RB fit runs in right occipital cortex. For this analysis, the statistical threshold was $\underline{t}(30) > 4$, in a contiguous cluster of greater than 350mm^3.

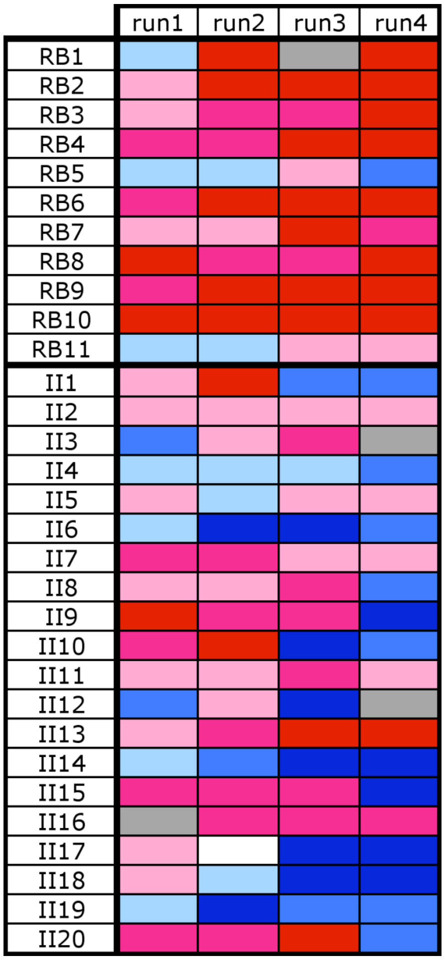
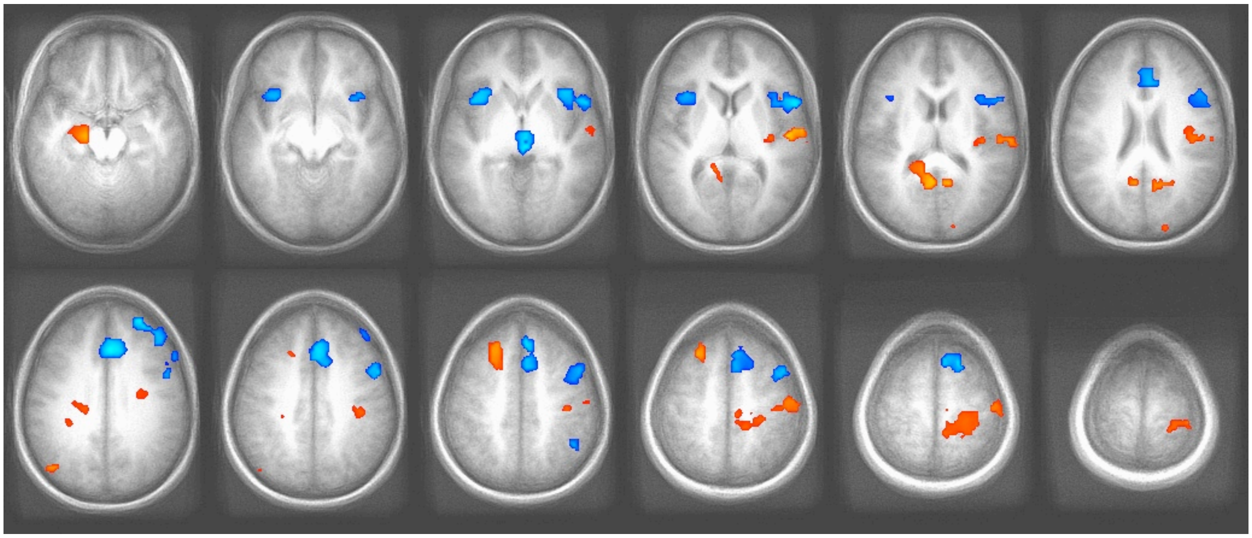


Figure 3.3: Exp. 2 model fits for RB and II subjects

Experiment 2: model fits for RB (top) and II subjects (bottom). Fit values were separated into thirds with increasing hue indicating better fit. RB fits are in warm and II fits in cool colors. The darkest hues were included in the fMRI analysis in Figure 3.4.

A)



B)

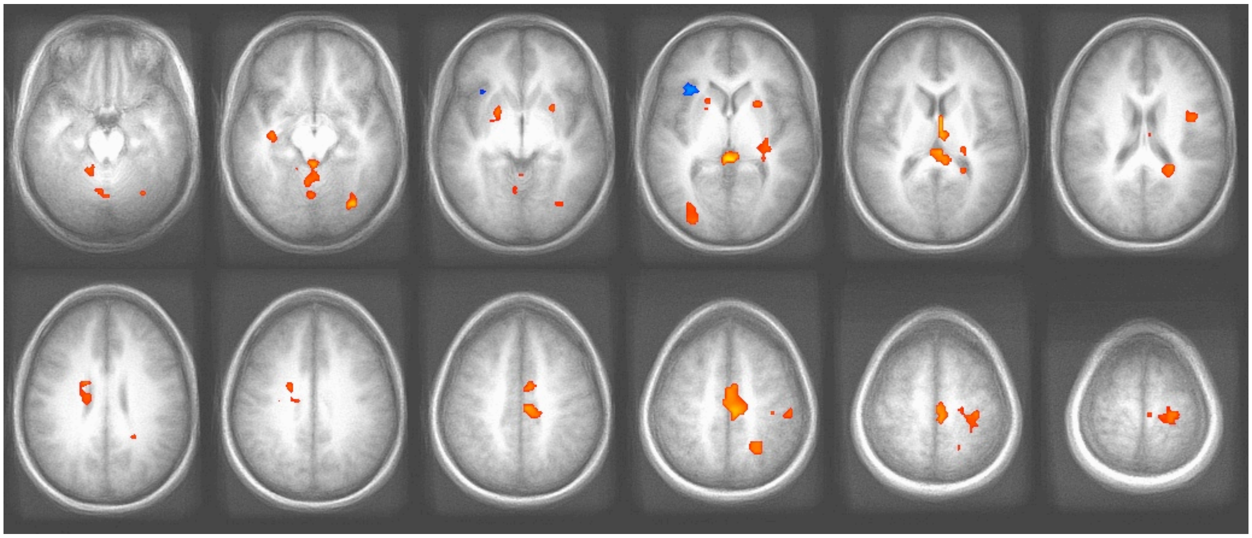


Figure 3.4: Exp. 2 model-identified best RB and II fMRI activity

Experiment 2: Correct vs. incorrect trial activity for (A) best RB model fit runs (C) best II model fit runs. All images were thresholded at T>3, cluster>350mm^3.

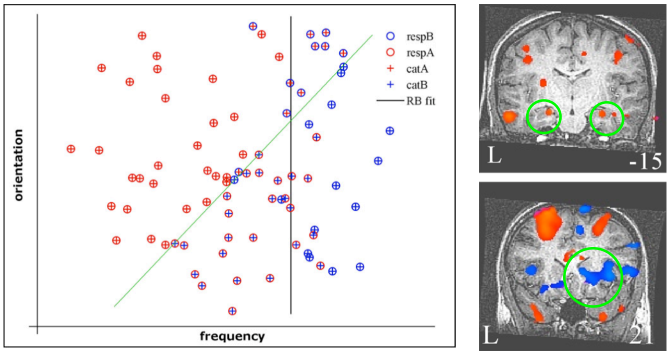
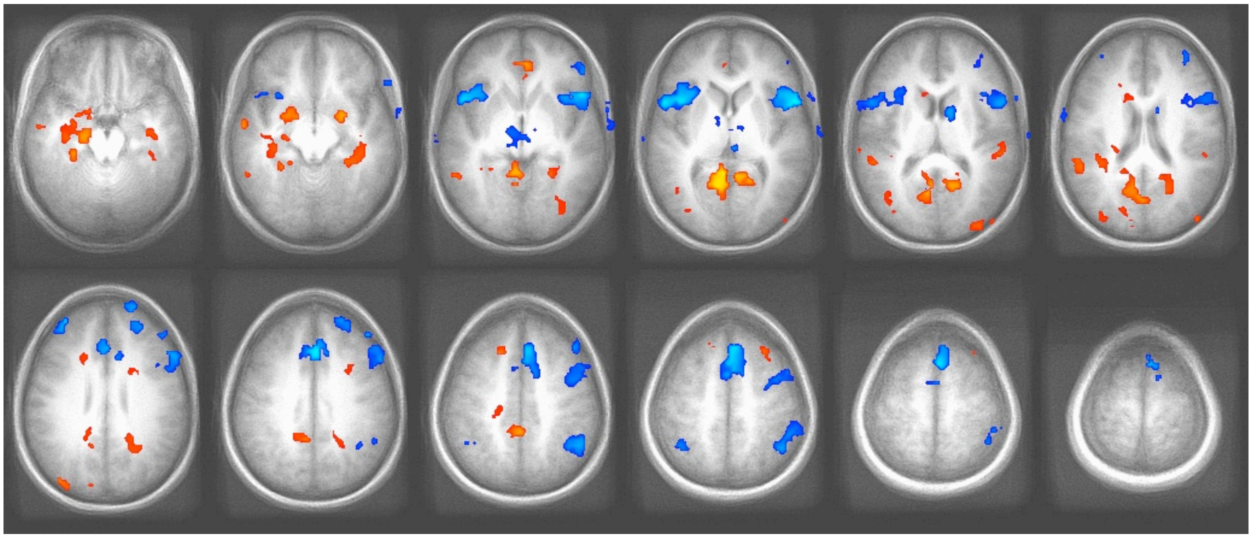


Figure 3.5: example of modeling in an individual subject

One II subject run that was best fit with the RB model (left) and the corresponding contrast of correct vs. incorrect trial activity (right). Plus signs indicate the category membership of the stimuli whereas the circles show the pattern of subject responses. Instances where the colors agree indicate correct responses and all others are incorrect. The dotted line indicates the optimal decision boundary and the solid line the best fitting account of the data according to the RB model. The corresponding functional activity resembles that of the best fit RB run group activity.

A)



B)

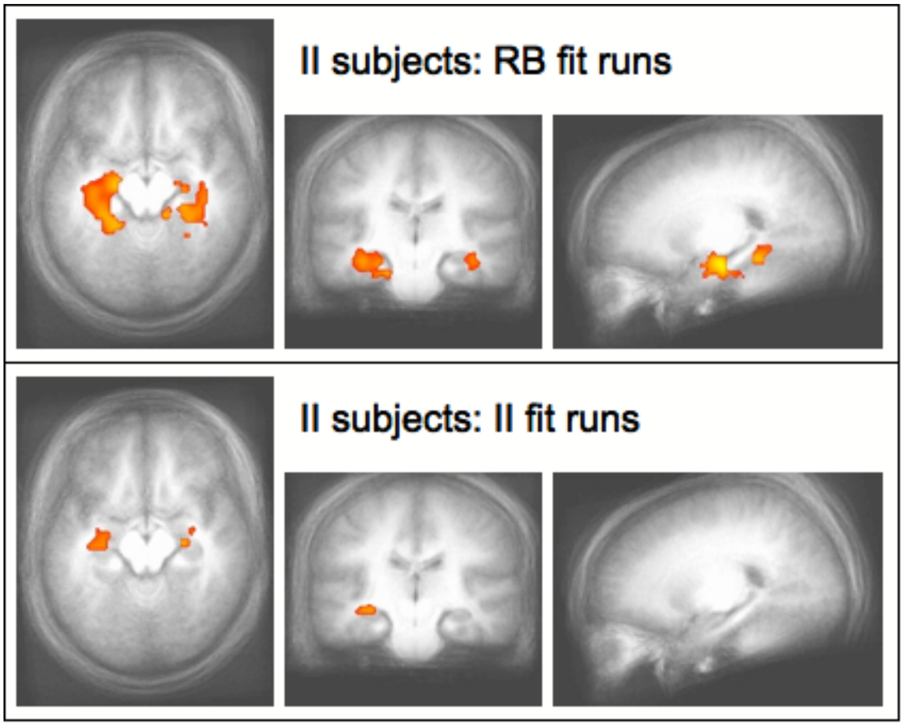


Figure 3.6: best fitting RB runs in II subjects

(A) Similar to the analysis in Figure 3.5, correct vs. incorrect trial activity is contrasted here from the best fitting RB runs that occurred within II subjects. (B) MTL ROI analysis of II subject correct vs. incorrect trial activity for RB fit runs (top) and II fit runs (bottom).

CHAPTER 4 FIGURES

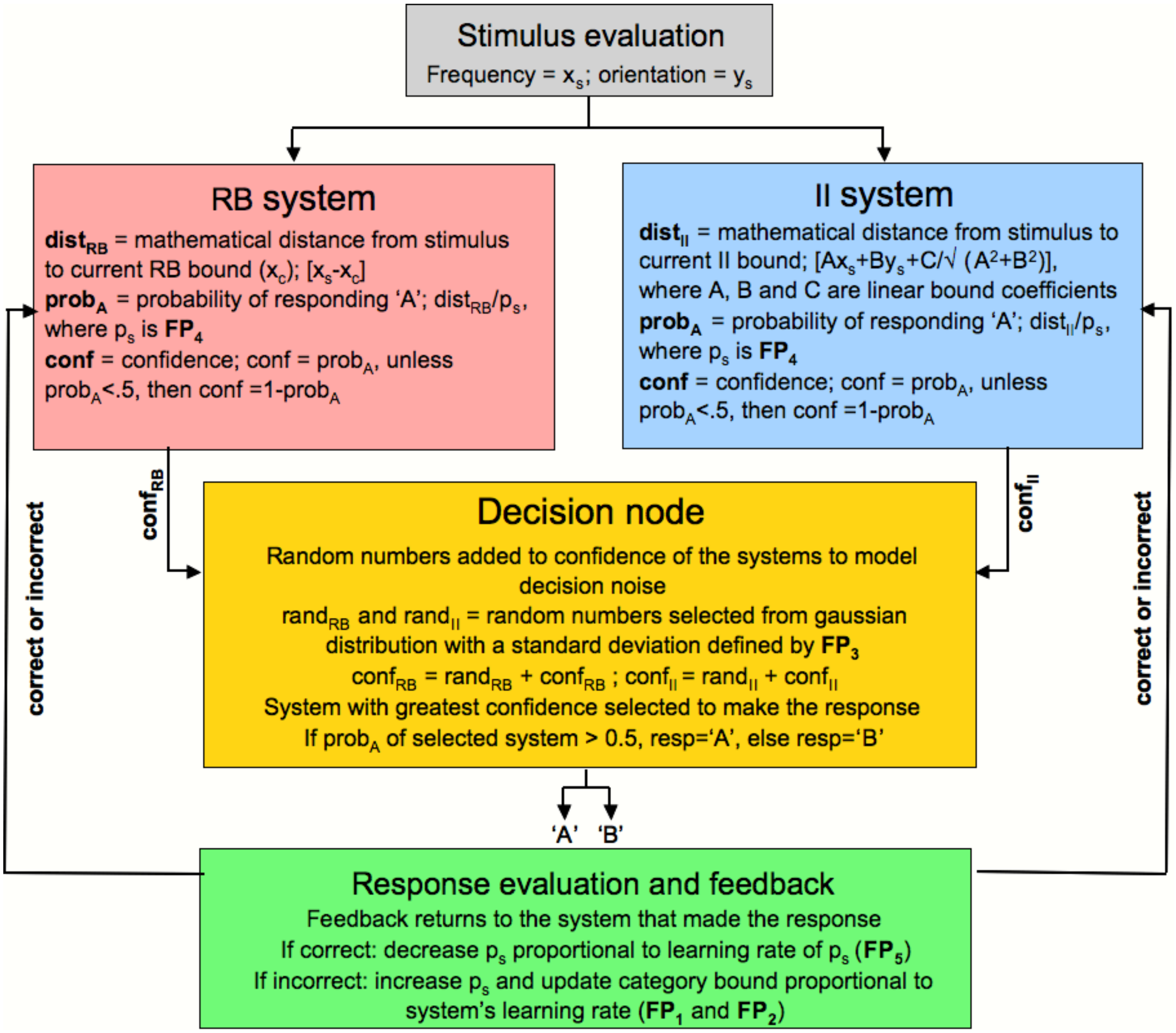


Figure 4.1: Schematic structure of PINNACLE

Schematic of the PINNACLE computational model and the accompanying free parameters. Stimulus information is fed into the RB and II systems and evaluated separately. A categorical decision is made within each system, but the decision node adjudicates between the systems according to their relative confidence levels. After a response is made, feedback returns to the system that made the response, and in the case of negative feedback, the representation is updated.

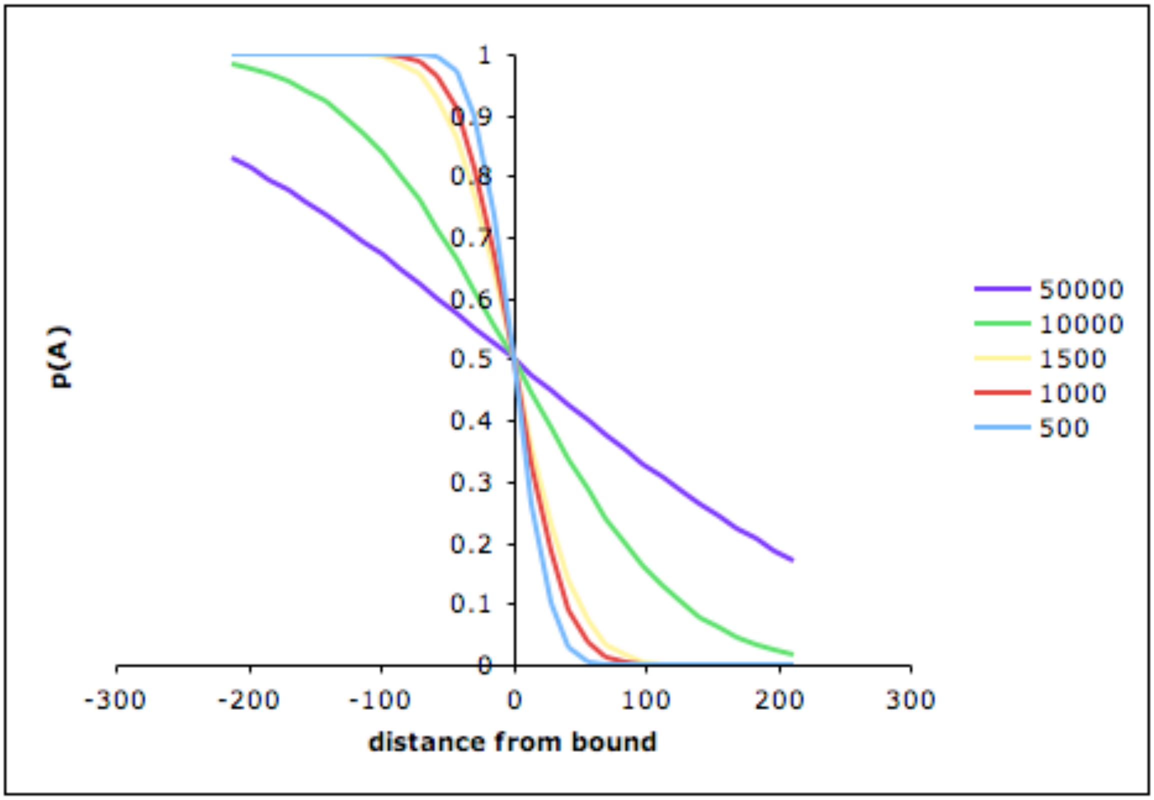


Figure 4.2: perceptual shaping parameter

Effect of the distance to the boundary on p(A) for different values of the perceptual shaping parameter. A higher perceptual shaping parameter means that the discrimination between the categories becomes more distinct, as reflected in the large degree of difference between p(A) values as a function of distance.

A)

	RUN1	RUN2	RUN3	RUN4
RB model	63.31	71.45	76.49	81.19
RB subj	64.39	73.74	79.82	81.16
II model	65.62	74.03	77.09	75.53
II subj	64.17	68.17	75.43	76.06

B)

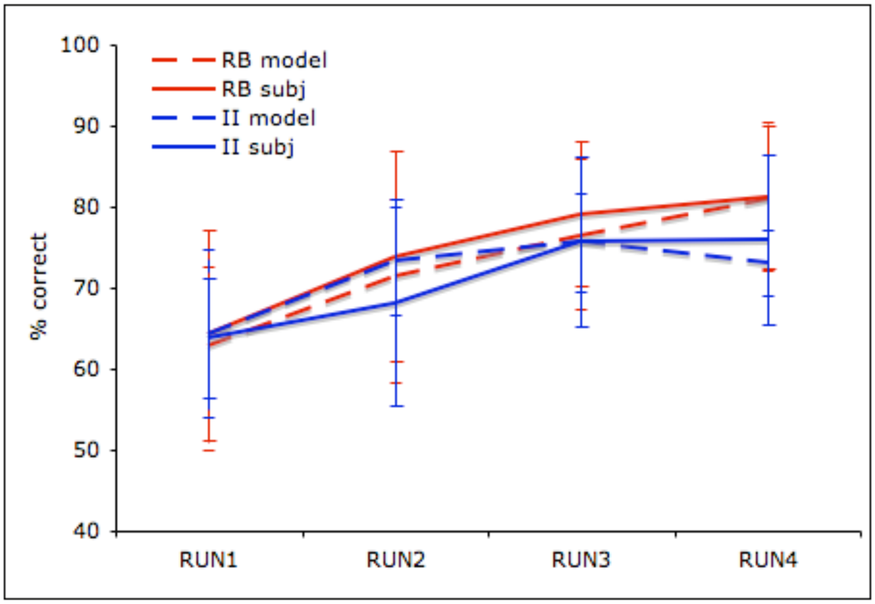


Figure 4.3: 4 parameter model fit

(A) Table of the accuracy of the model and the subjects on each run. (B) Best fitting 4 parameter model out of 12 different feedback models. Model accuracy versus RB and II subject group accuracy. Feedback model = 6. Fit value = 57.16

A)

	RUN1	RUN2	RUN3	RUN4
RB model	65.89	76.05	79.59	80.26
RB subj	64.39	73.74	79.82	81.16
II model	65.54	69.42	73.81	76.01
II subj	64.17	68.17	75.43	76.06

B)

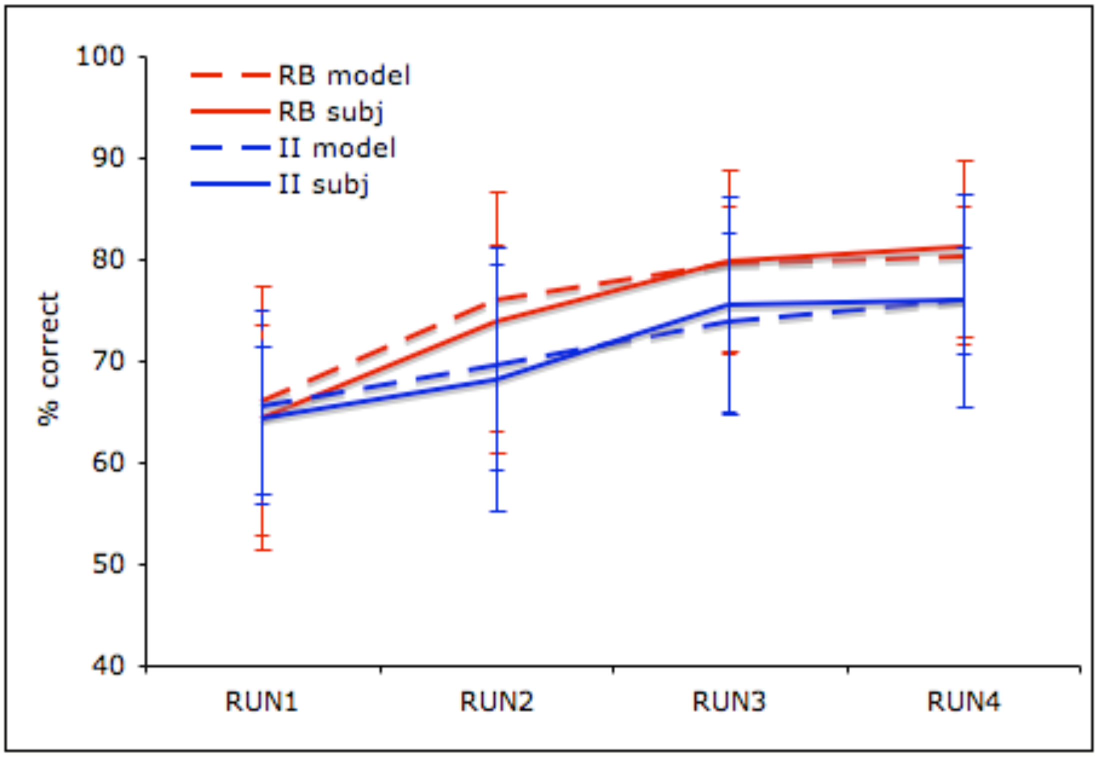


Figure 4.4: 5 parameter model fit

(A) Table of the accuracy of the model and the subjects on each run. (B) Best fitting 5 parameter model out of 12 different feedback models. Model accuracy versus RB and II subject group accuracy. Feedback model = 2. Fit value = 14.47

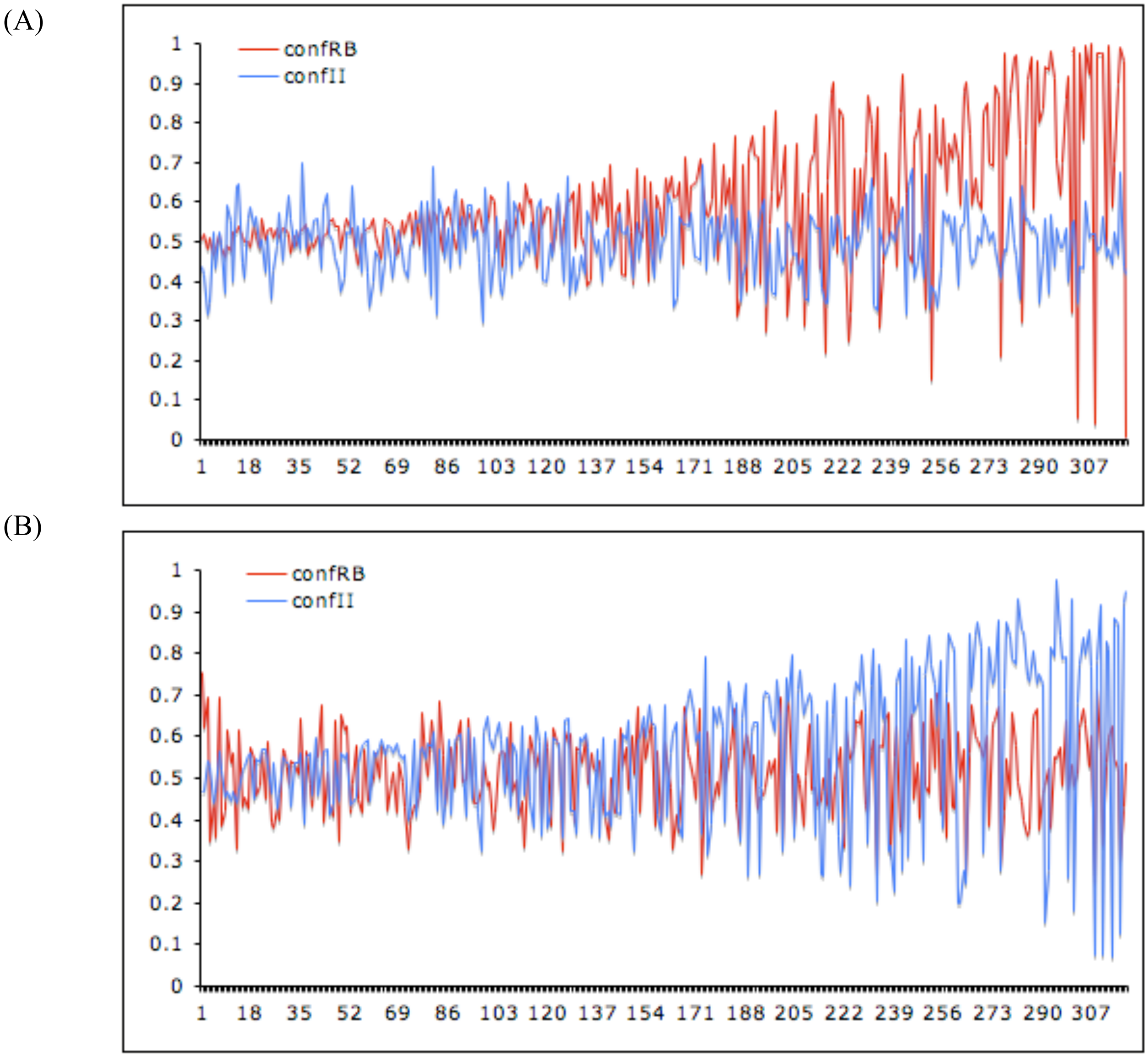


Figure 4.5: Individual subject PINNACLE trial fits

RB and II system confidence across 320 trials in an (A) RB subject and (B) II subject.

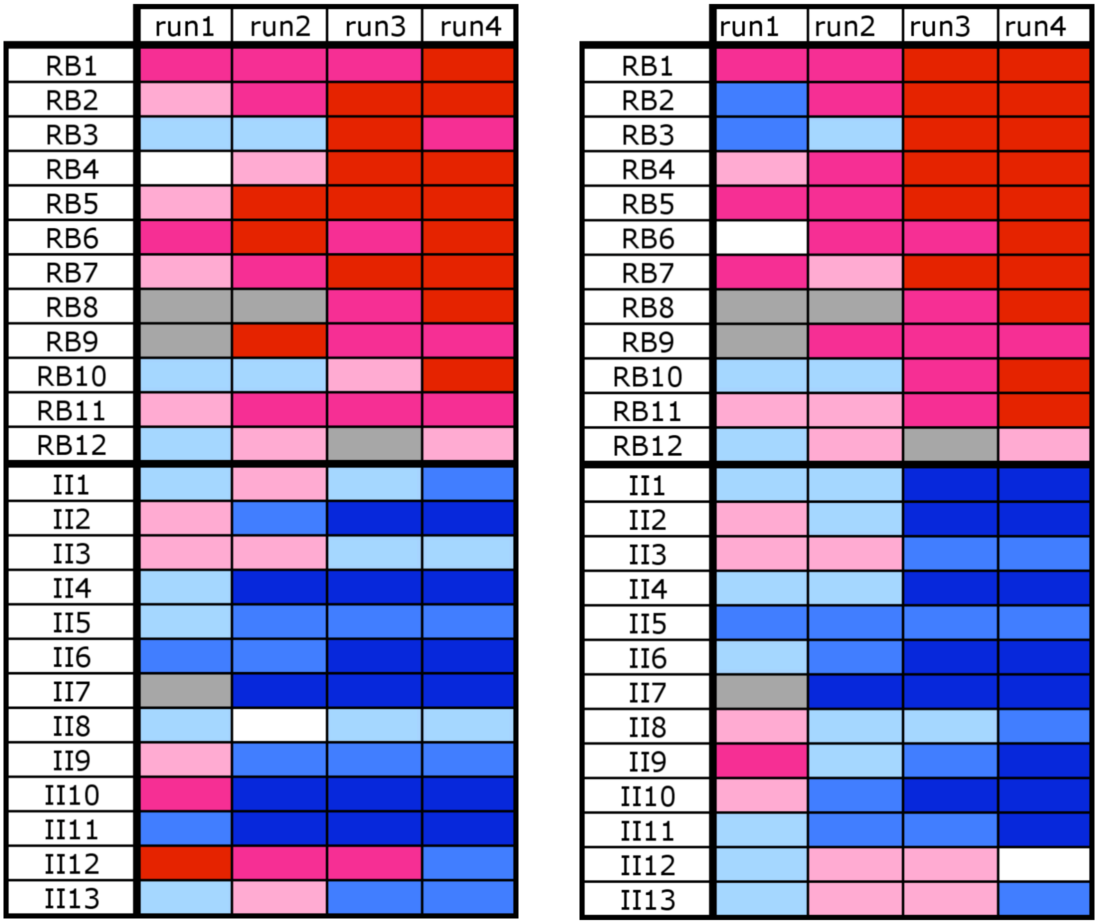


Figure 4.6: Exp. 1 avg RB and II fit trials

Experiment 1. (A) Block model run characterization (B) PINNACLE system engagement per block. Of note here is the similarity between the mathematical models and PINNACLE on a block level.

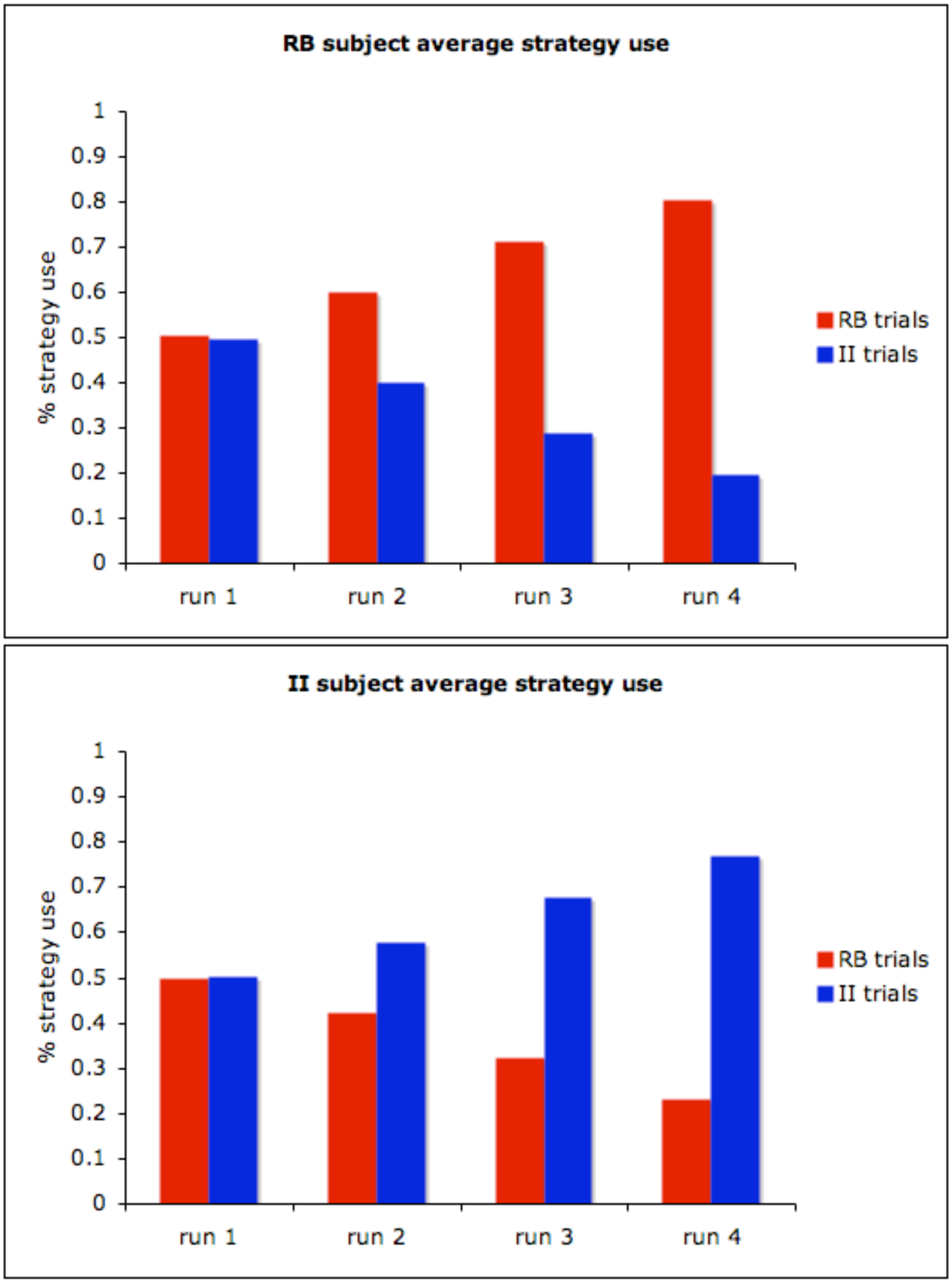


Figure 4.7: Exp. 1 avg RB and II fit trials

Experiment 1. (A) RB and (B) II subject group average strategy use per block. In general RB trials dominate RB subjects and II trials dominate II subjects.

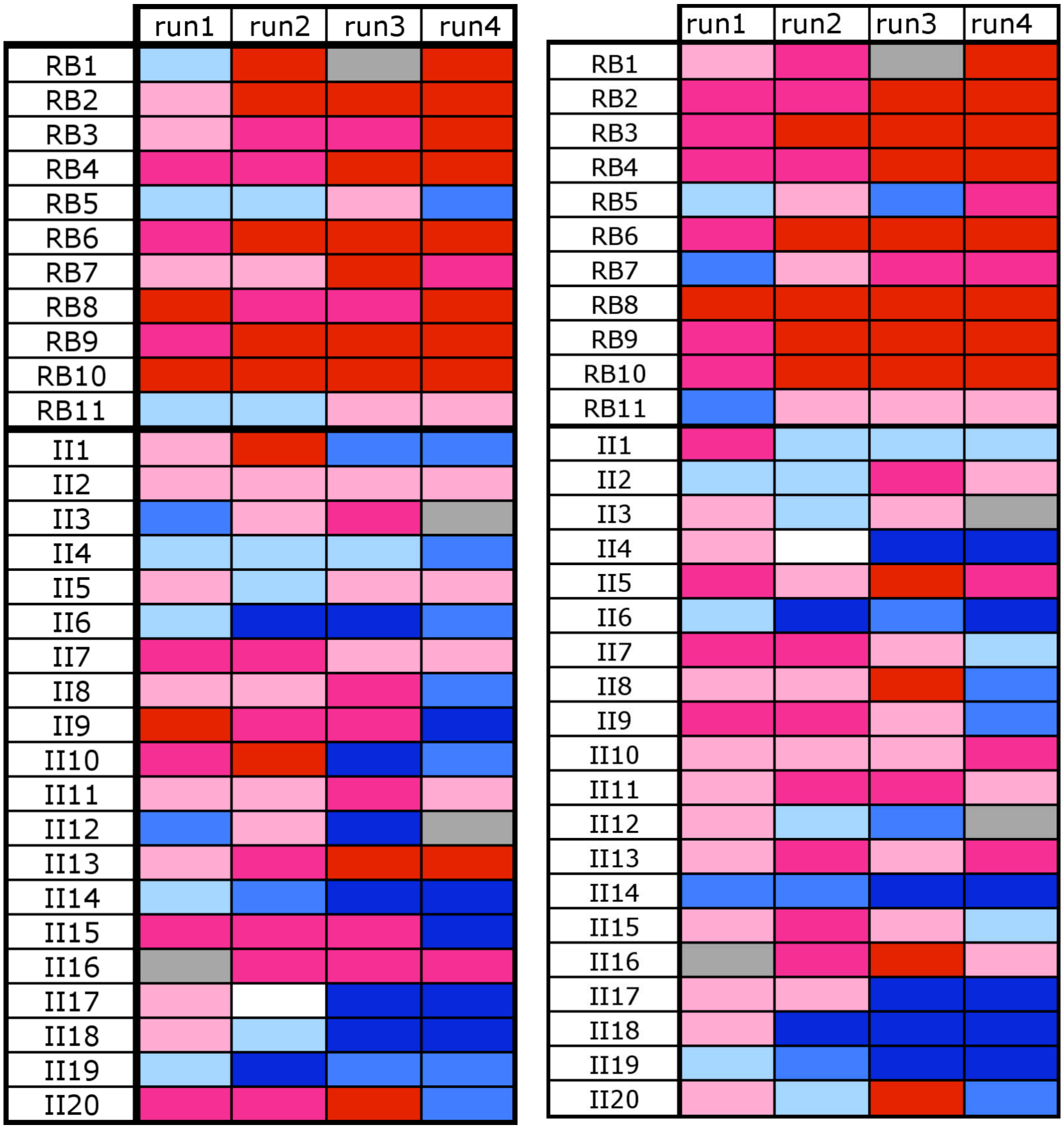


Figure 4.8: Exp. 2 avg RB and II fit trials

Experiment 2. (A) Block model run characterization (B) PINNACLE strategy predictions averaged over 80-trial blocks. Of note here is the similarity between the mathematical models and PINNACLE on a block level.

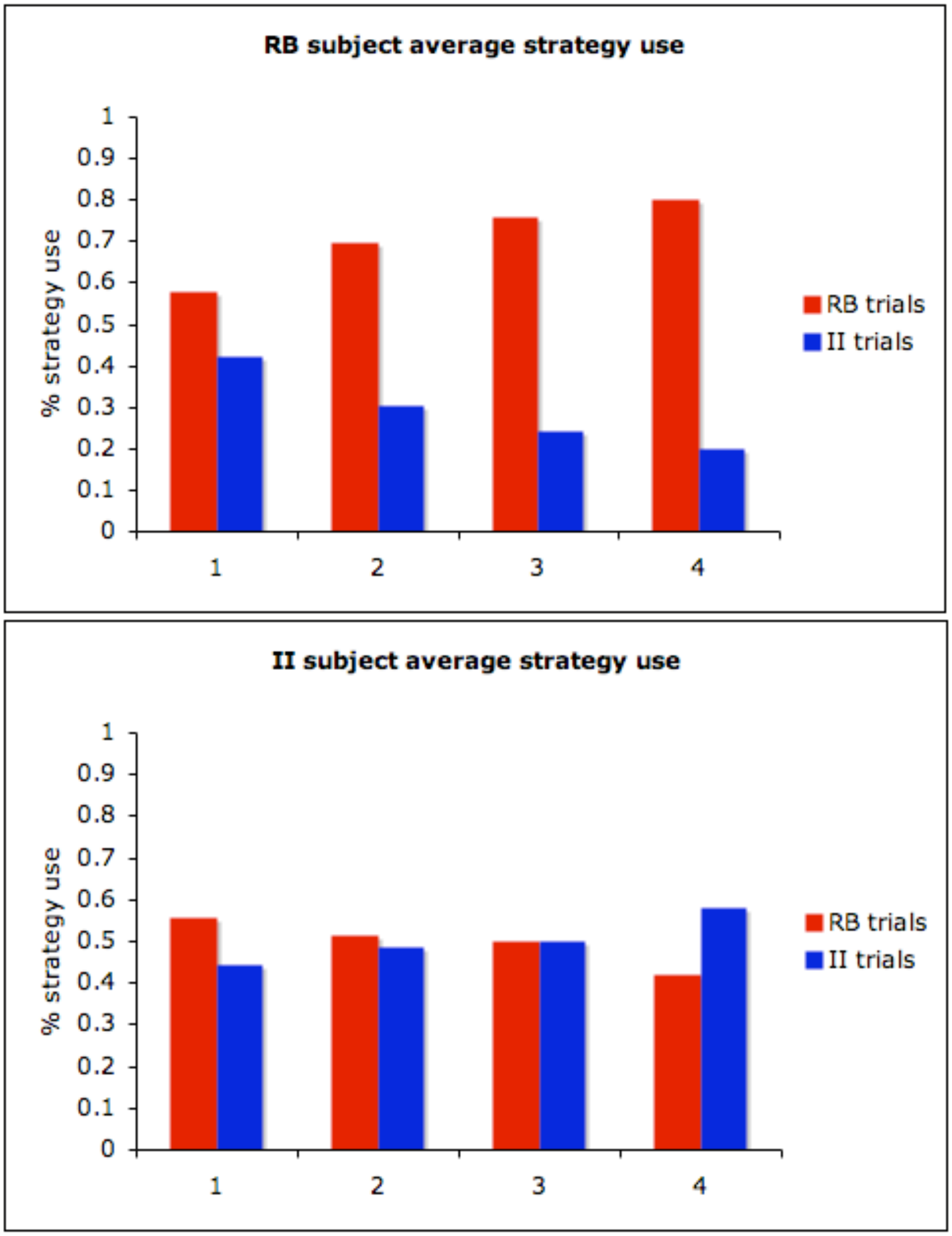


Figure 4.9: Exp. 2 avg RB and II fit trials

Experiment 2. (A) RB and (B) II subject group average strategy-use per block according to PINNACLE. Compared to Experiment 1, the II group in Experiment 2 more frequently appears to adopt a mixture of strategies.

Table 4.1: Assumptions of the PINNACLE model

Assumptions made in the construction of the PINNACLE model.

1	There are only 2 systems: RB and II
2	The RB system is subserved by an explicit memory process (MTL) in concert with working memory and attention (PFC and ACC)
3	The II system relies on perceptual learning within the tail of the caudate in communication with cortical visual association areas
4	The systems interact competitively (rather than cooperatively or in parallel with no interaction)
5	Feedback only returns to the system that made the response.
6	The category representation is only adjusted on incorrect trials.
7	Competition is resolved by the PFC through top-down modulation of both the RB and II systems

Table 4.2: Different feedback mechanisms for the PINNACLE model

Model Number	Description
1	If the RB system is selected, the RB system gets feedback. If the II system is selected, the II system gets the feedback. This is true on both correct and incorrect trials.
2	Both systems get feedback on every trial.
3	Feedback only occurs on correct trials and goes to both systems.
4	Both systems get feedback on every incorrect trial.
5	If the RB system is selected, both systems get feedback. If the II system is selected, only the II system gets feedback.
6	If the II system is selected, both systems get feedback. If the RB system is selected, only the RB system gets feedback.
7	If the RB system is selected on an incorrect trial, both systems get feedback. Only the II system gets feedback on every other trial.
8	If the II system is selected on an incorrect trial, both systems get feedback. Only the RB system gets feedback on every other trial.
9	If the RB system is selected on a correct trial, it receives feedback. If the II system is selected on an incorrect trial it receives feedback.
10	If the II system is selected on a correct trial, it receives feedback. If the RB system is selected on an incorrect trial, it receives feedback.
11	Feedback only occurs on correct trials and goes to the system that was assigned to the response.
12	Feedback only occurs on incorrect trials and goes to the system that was assigned to the response.

Table 4.3: Comparison of feedback models (A) 4 parameter model fits (B) 5 parameter model fits

In the following tables, the data is sorted according to fit value, with increasing values indicating worse fits to the data. The version of the feedback model corresponds to the model number in Table 4.2. The result of the minimization procedure was a set of parameters: lrmodel (learning rate of the model; for the 4-parameter model), lrRB (learning rate of the RB system; for the 5-parameter model), lrII (learning rate of the II system; for the 5-parameter model), lrPS (learning rate of the perceptual shaping parameter), PS (starting perceptual shaping parameter) and Gauss sigma (standard deviation of the Gaussian noise distribution parameter).

A)

Feedback	Fit value	lrmodel	lrPS	PS	Gauss sigma
6	57.1645	0.0409	0.0321	455	16.7240
3	89.6892	0.4507	0.4062	236791	18.6294
2	131.2210	0.0469	0.1295	52565	7.4734
8	138.3309	0.0317	0.0829	251	4.8517
11	138.9801	3.4206	0.4750	506774	9.8050
4	182.1173	0.1098	0.0994	5571	0.3713
1	189.5899	0.0702	0.2123	45444	17.9382
12	260.5065	0.5638	0.1321	11906	0.9196
9	295.0172	0.5714	0.1681	110351	0.4685
7	342.8350	0.0385	0.0033	1057	6.6598
5	453.7860	0.0341	0.0553	6128	15.1179
10	1141.1859	0.3371	0.1887	150762	34.4861

B)

Feedback	Fit value	lrRB	lrII	lrPS	PS	Gauss sigma
2	14.6242	0.1703	0.0288	0.5382	145932	1.2043
1	29.0467	0.5203	0.0357	0.3689	99857	7.1001
6	44.4372	0.0601	0.3661	0.5796	156153	15.5632
3	50.3389	0.9612	0.3996	0.3906	172913	3.4329
5	84.9332	0.5259	0.0169	0.2025	176041	7.2375
4	160.0943	0.8720	0.1061	0.1186	64529	0.2630
12	178.4442	0.7514	0.1754	0.2840	56990	6.6114
8	213.5469	0.0408	0.1129	0.3195	119288	31.2168
11	263.6949	0.9084	1.0447	0.3700	210546	18.4908
9	305.5924	0.7797	1.3193	0.4274	201687	43.6479
7	809.5260	1.5725	0.0318	0.5911	163536	20.0740
10	1102.3731	0.9352	0.4984	0.4274	209661	38.9850

CHAPTER 5 FIGURES

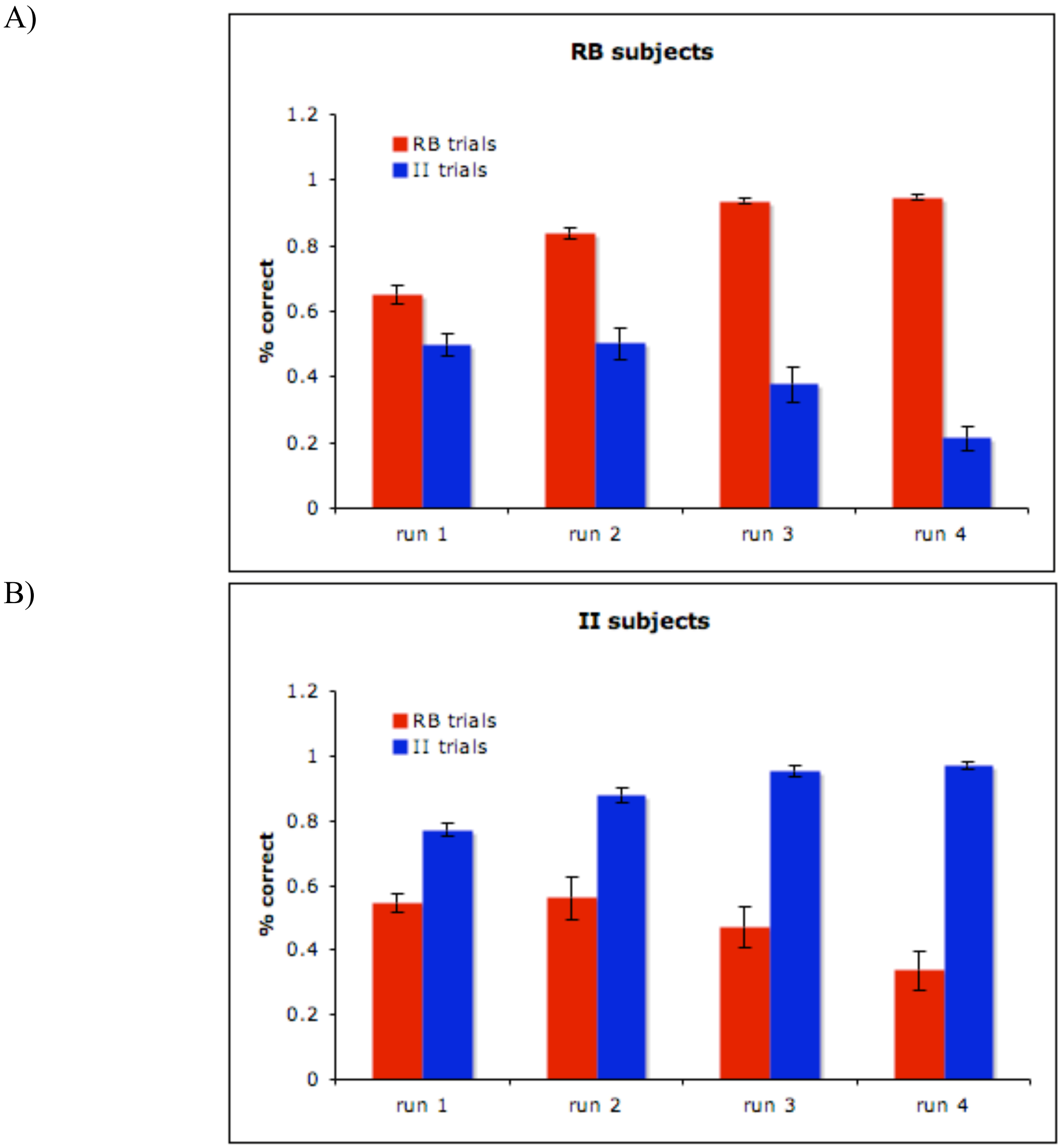


Figure 5.1: Exp. 1 PINNACLE trial accuracy in (A) RB subjects and (B) II subjects

PINNACLE designated every trial according to whether the RB or II system was active. When trials were grouped according to this designation between RB and II subjects there were differences in accuracy. RB trials were more accurate in RB subjects and II trials were more accurate in II subjects.

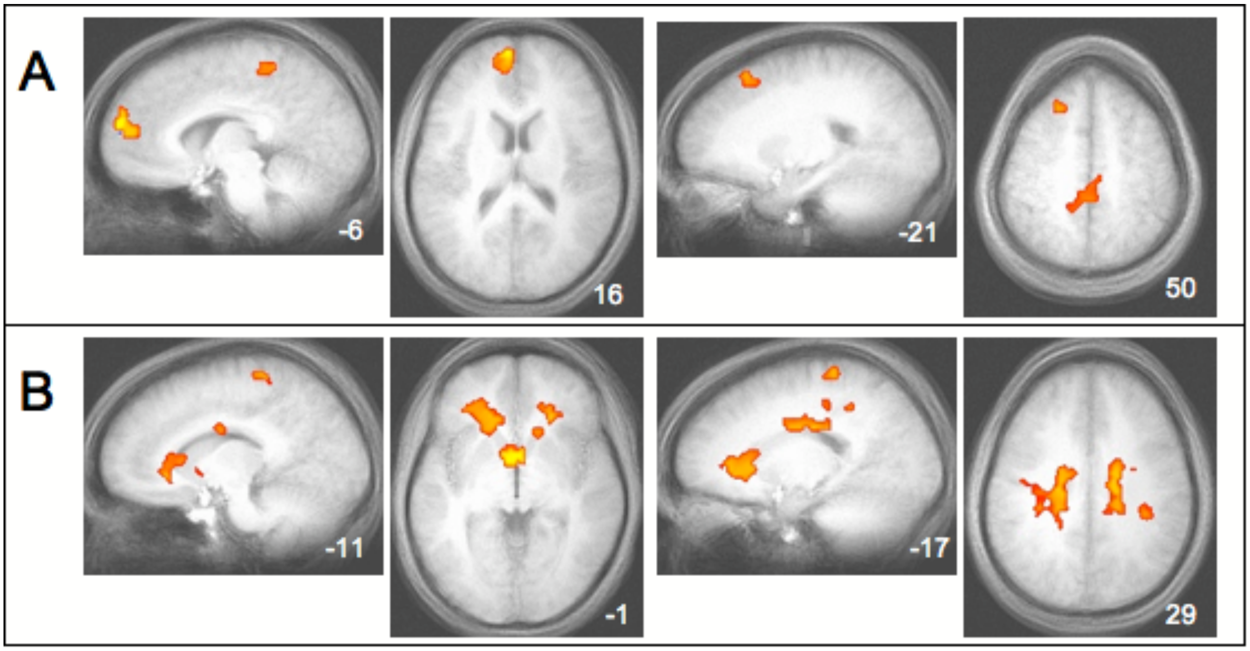


Figure 5.2: Exp. 1 Correct – incorrect RB fit trials in (A) RB subjects and (B) II subjects

PINNACLE-based trial sorting reveals whole-brain activity associated with categorization success on RB trials. This activity differs between the RB and II subject groups, with one common area of activity in the parietal lobe.

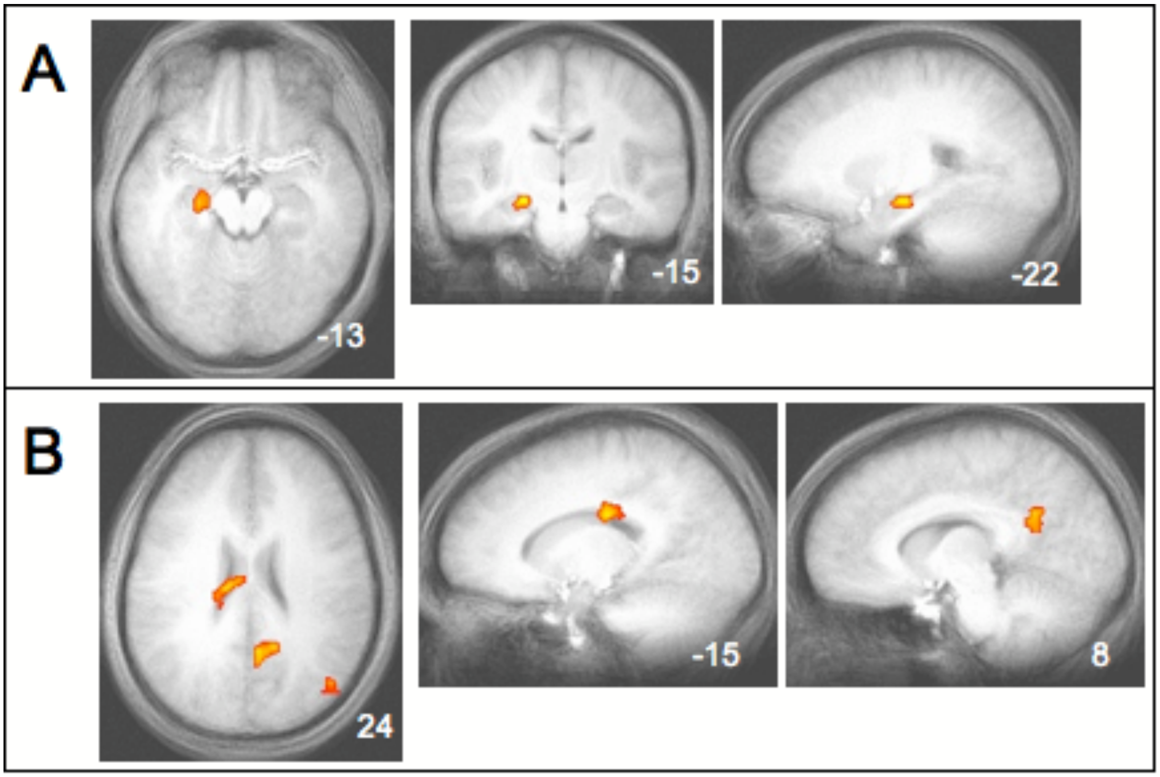


Figure 5.3: Exp. 1 Correct – incorrect II fit trials in (A) RB subjects and (B) II subjects

PINNACLE-based trial sorting reveals whole-brain activity associated with categorization success on II trials. This activity differs between the RB and II subject groups.

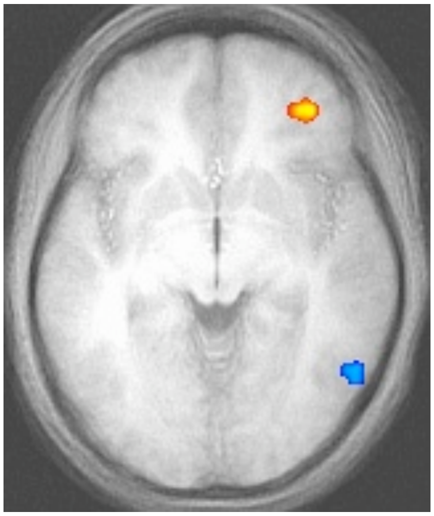


Figure 5.4: Exp. 1 Best fitting RB blocks versus best fitting II blocks in correct trials.

PINNACLE-based trial sorting reveals whole-brain activity associated with best fitting RB blocks compared to that during best fitting II blocks. This analysis is restricted to activity associated with correct trials as these best fitting blocks contain few incorrect trials. A region in right PFC is more active for best fitting RB blocks and a region in right posterior visual association cortex is more active for best fitting II blocks.

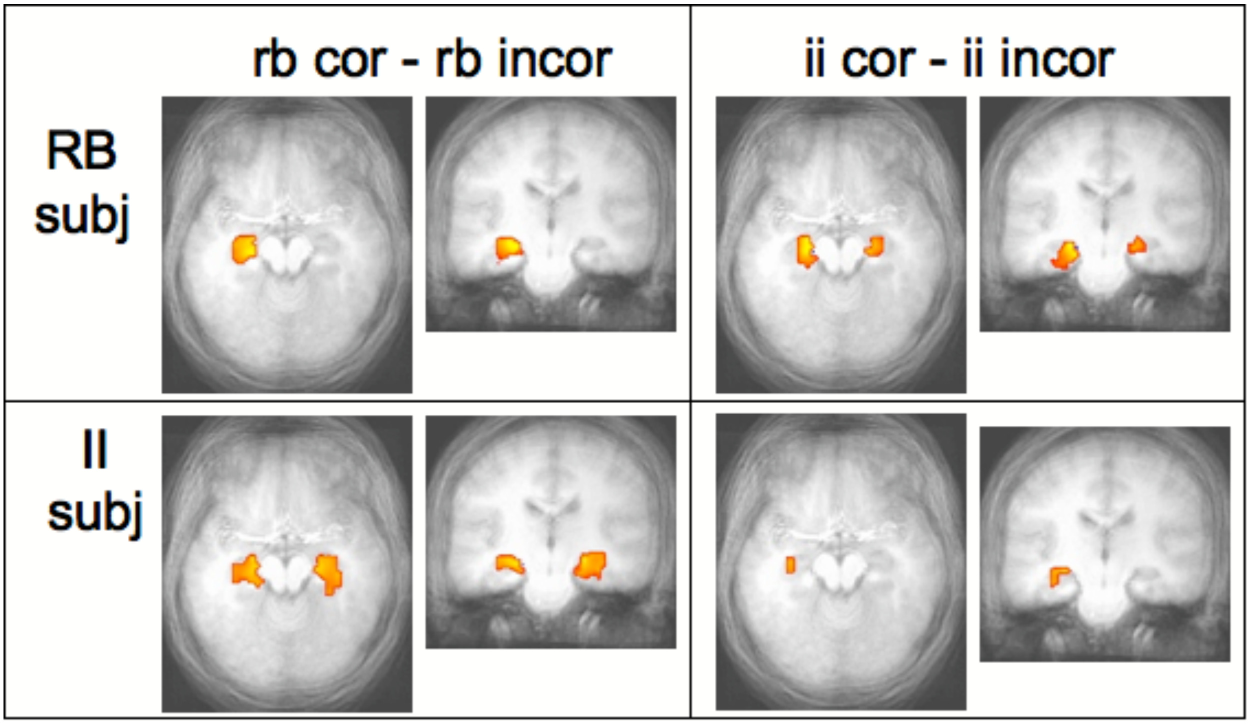


Figure 5.5: Exp. 1 MTL ROI RB and II fit trials

Within the MTL ROI, PINNACLE-based trial sorting reveals differences in successful categorization activity for RB and II trials in both the RB and II subject groups. The most robust differences occur within both subject groups in PINNACLE-identified RB trials.

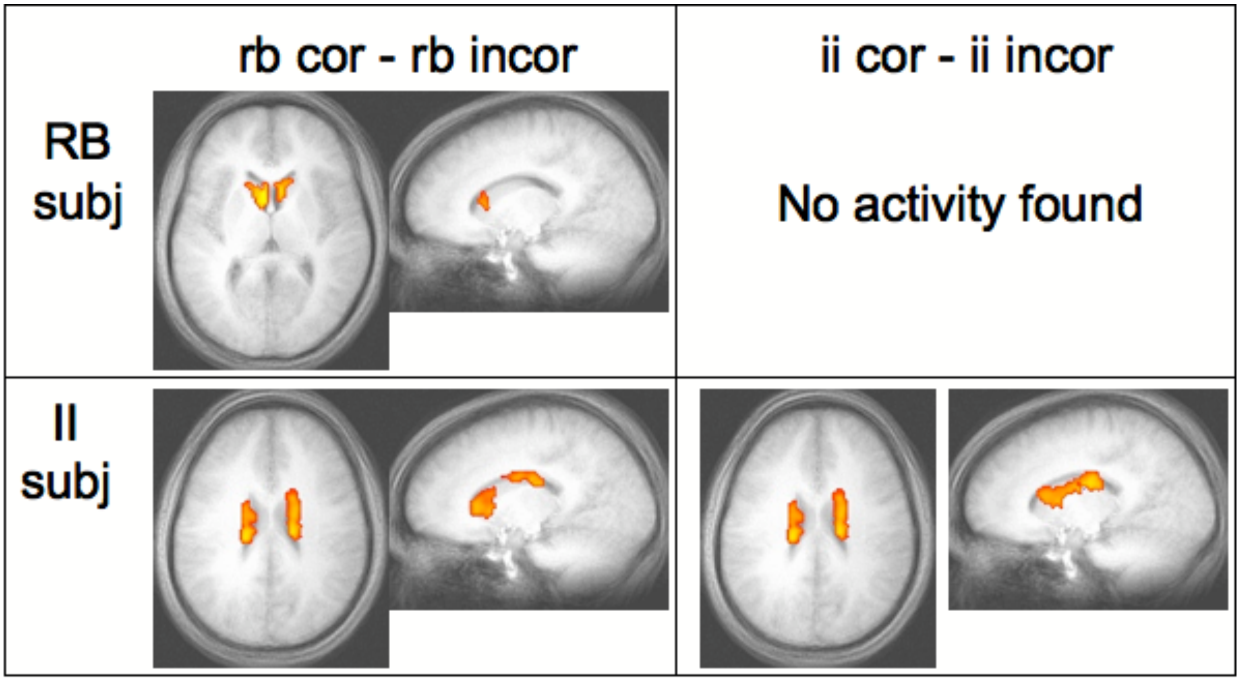


Figure 5.6: Exp. 1 Caudate ROI RB and II fit trials

Within the caudate ROI, PINNACLE-based trial sorting reveals differences in successful categorization activity for RB and II trials in both the RB and II subject groups. During RB trials the II subject group shows bilateral caudate activity but the RB group only shows head of the caudate activation. Only the II subject groups shows bilateral caudate activity during II trials.

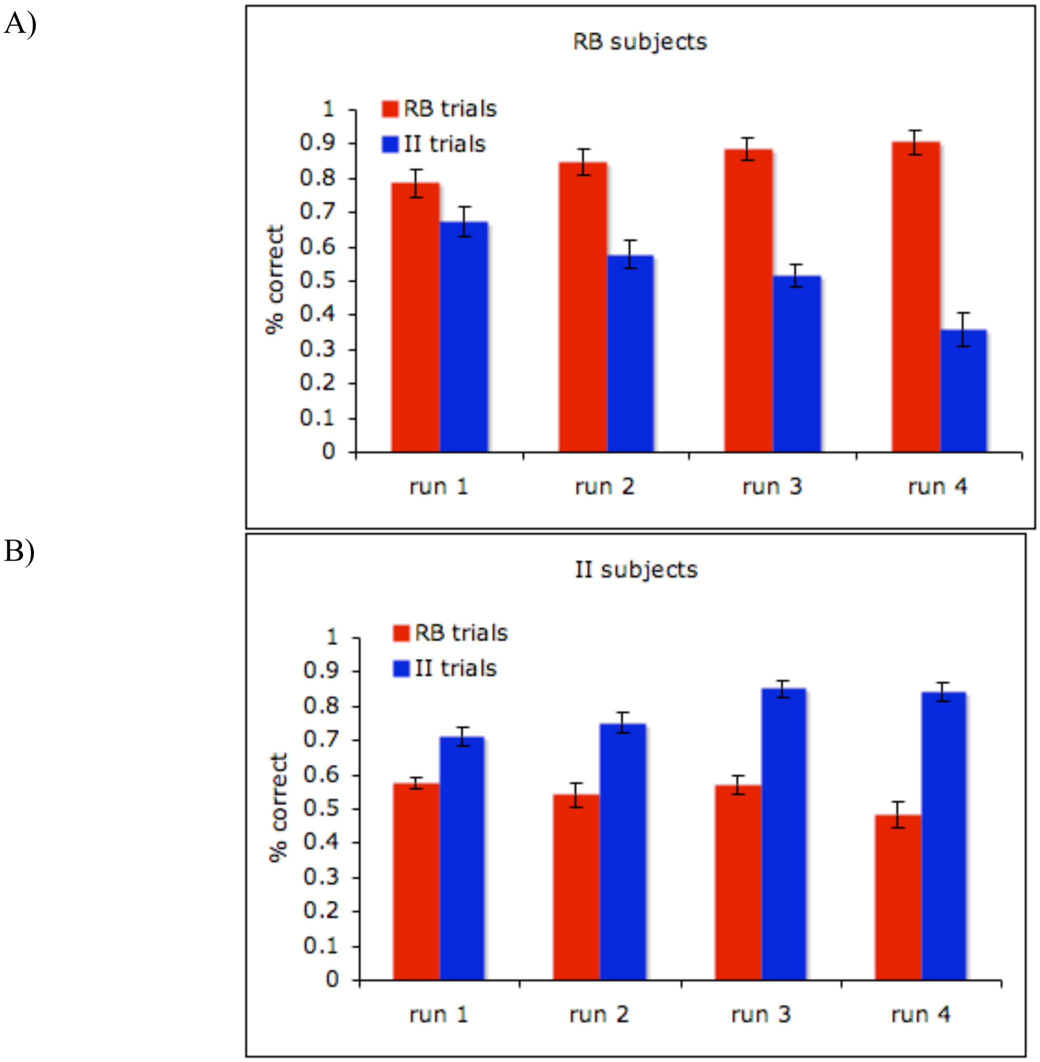


Figure 5.7: Exp. 2 PINNACLE accuracy in (A) RB and (B) II subjects.

PINNACLE designated every trial according to whether the RB or II system was active. When trials were grouped according to this designation between RB and II subjects there were differences in accuracy. RB trials were more accurate in RB subjects and II trials were more accurate in II subjects.

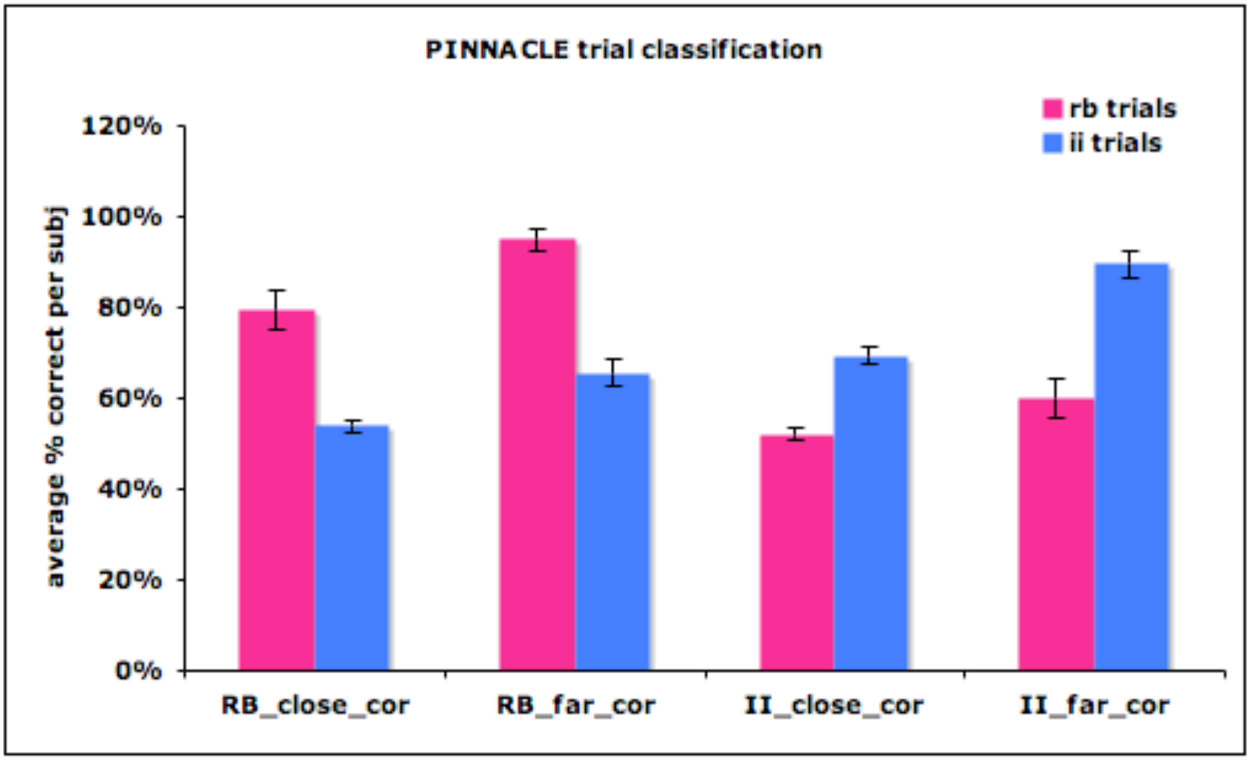


Figure 5.8: Exp. 2 Accuracy vs. Distance within RB and II fit trials in RB subjects

The increased variance in the stimuli in Experiment 2 allowed for a separation of close and far trial types. Within these sub-groups RB and II trials showed a similar pattern that far trials were more accurate than close trials in both RB and II subjects.

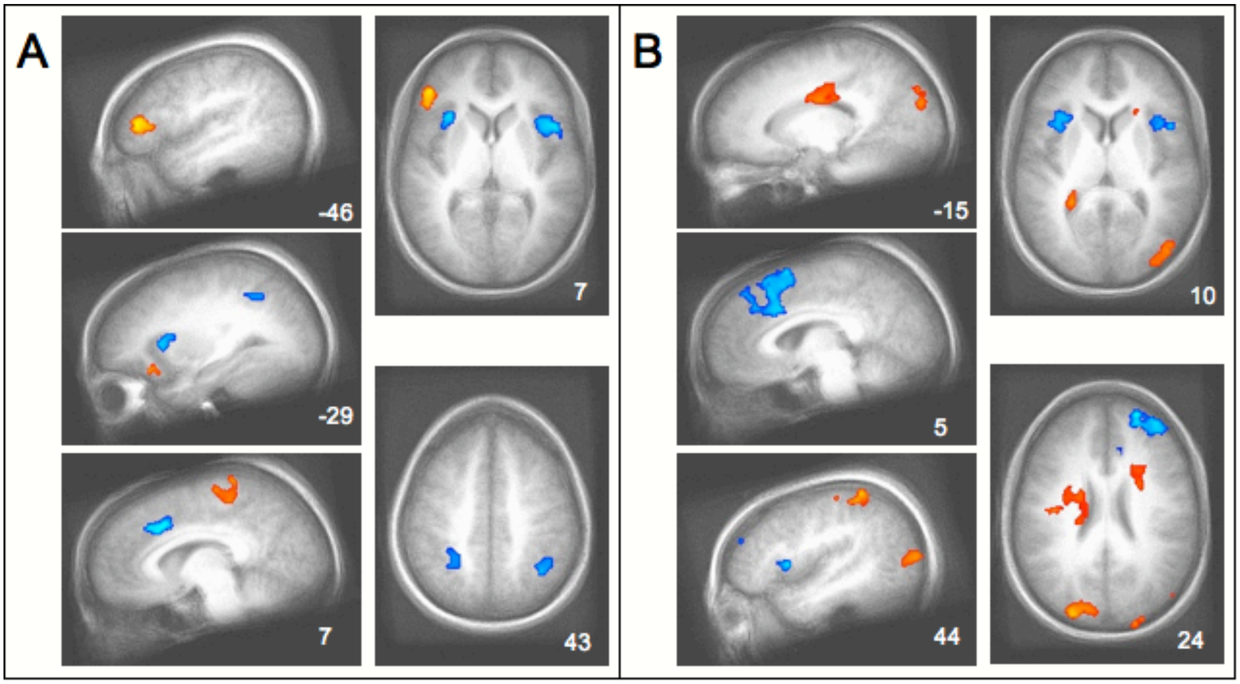


Figure 5.9: Exp. 2 Correct - incorrect RB fit trials in (A) RB and (B) II subjects.

PINNACLE-based trial sorting reveals whole-brain activity associated with categorization success on RB trials.

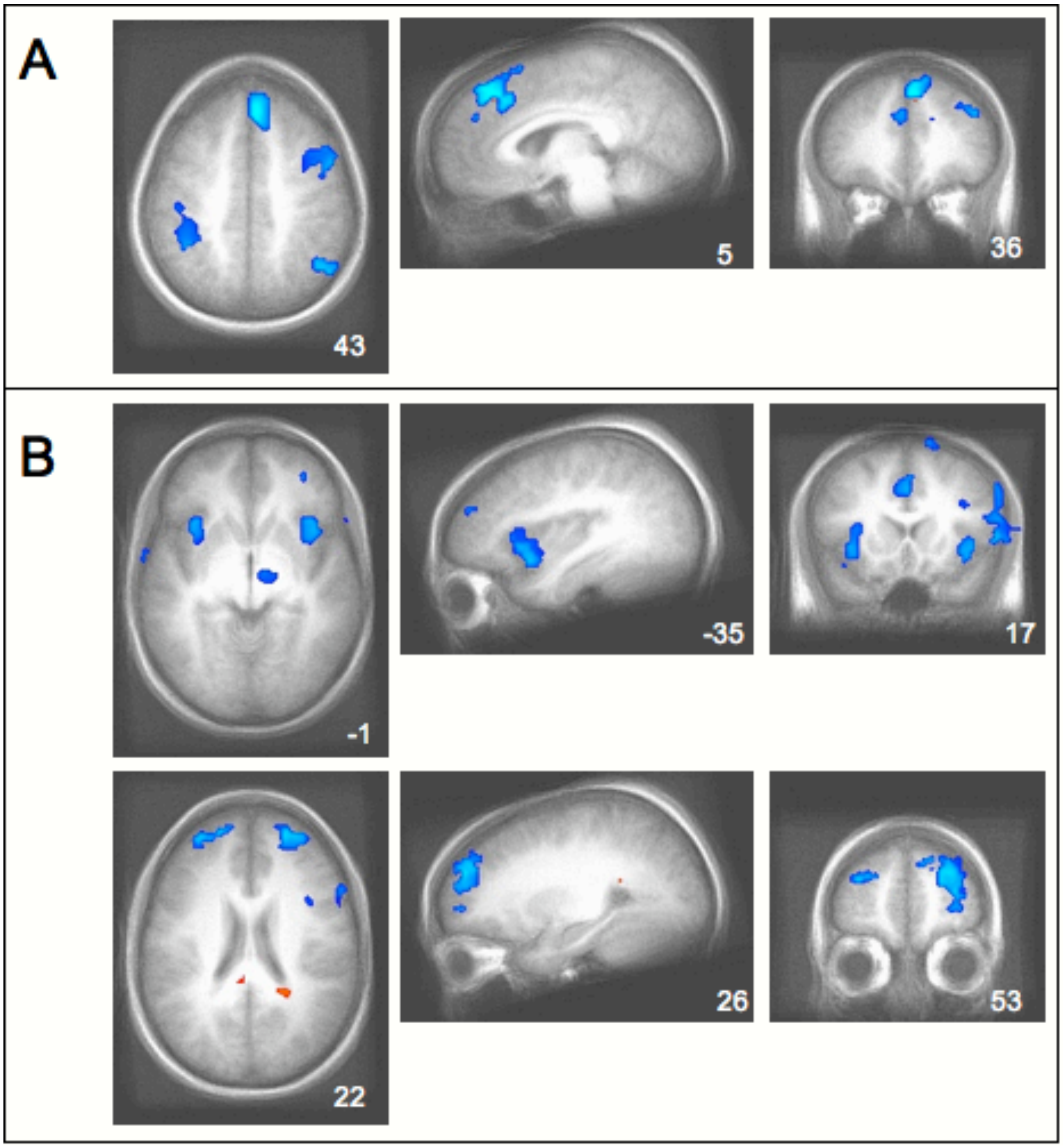


Figure 5.10: Exp. 2 Correct - incorrect II fit trials in (A) RB and (B) II subjects.
PINNACLE-based trial sorting reveals whole-brain activity associated with categorization success on II trials.

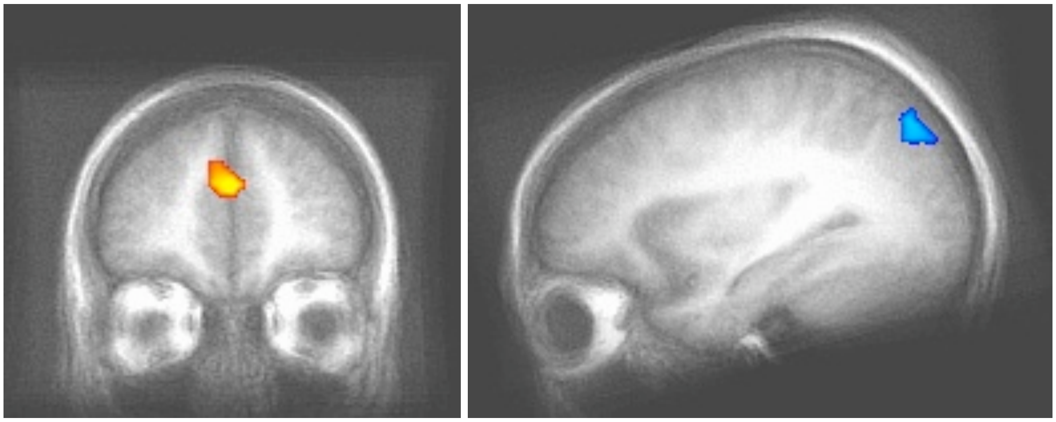


Figure 5.11: Exp. 2 Best fitting RB blocks versus best fitting II blocks in correct trials.

PINNACLE-based trial sorting reveals whole-brain activity associated with best fitting RB blocks compared to that during best fitting II blocks. This analysis is restricted to activity associated with correct trials as these best fitting blocks contain few incorrect trials. A region in medial PFC is more active for best fitting RB blocks (left) and a region in right posterior visual association cortex is more active for best fitting II blocks (right).

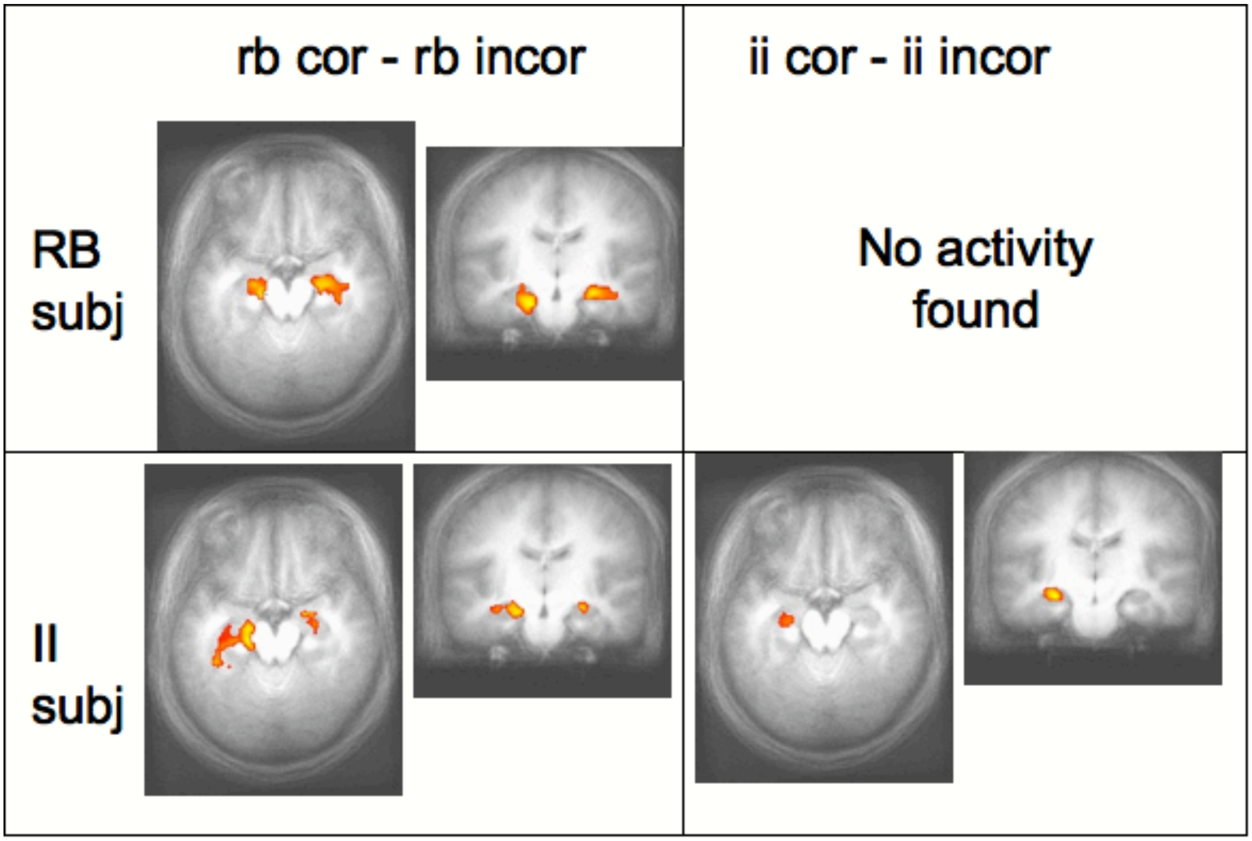


Figure 5.12: Exp. 2 MTL ROI RB and II fit trials

Within the MTL ROI, PINNACLE-based trial sorting reveals differences in successful categorization activity for RB and II trials in both the RB and II subject groups. The most robust differences occur within both subject groups in PINNACLE-identified RB trials. The II subjects also show left anterior MTL activity during II trials.

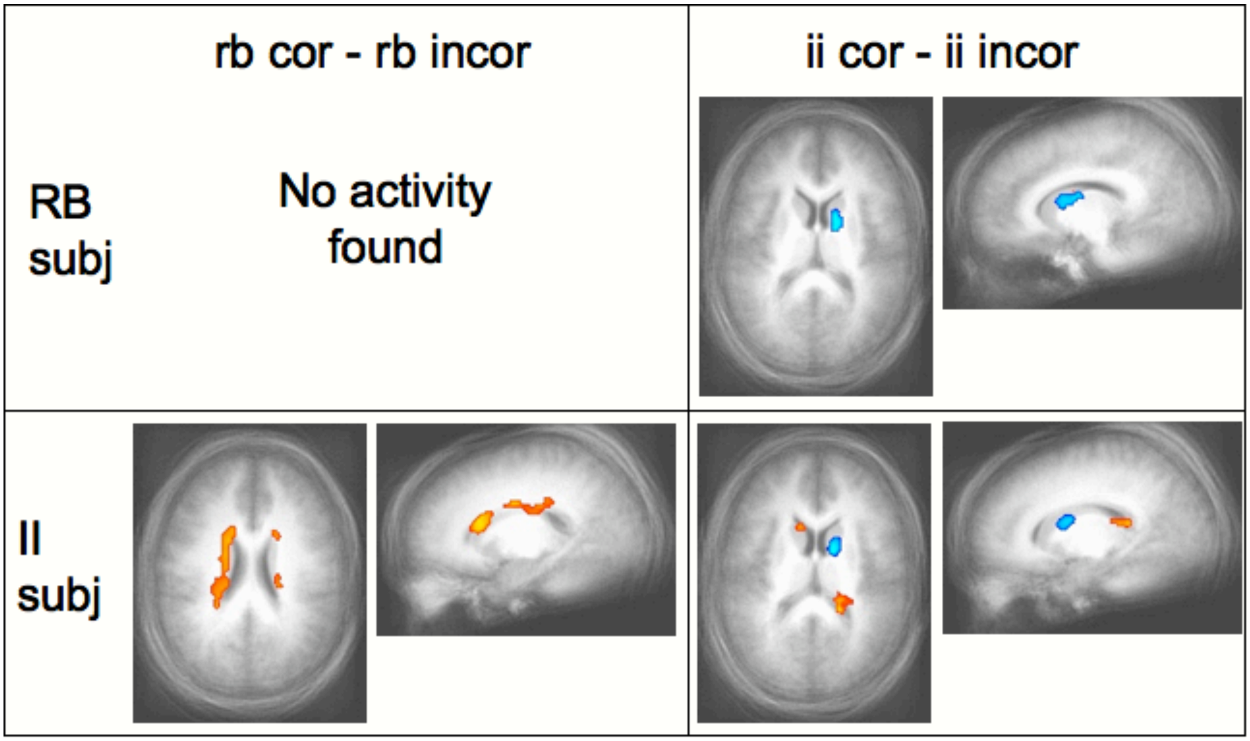


Figure 5.13: Exp. 2 Caudate ROI RB and II fit trials

Within the caudate ROI, PINNACLE-based trial sorting reveals differences in successful categorization activity for RB and II trials in both the RB and II subject groups. During RB trials only the II subject group shows bilateral caudate activity. Only the II subject group shows right posterior caudate activity during II trials. Both the RB and II groups have similar negative activity in anterior caudate during II trials.

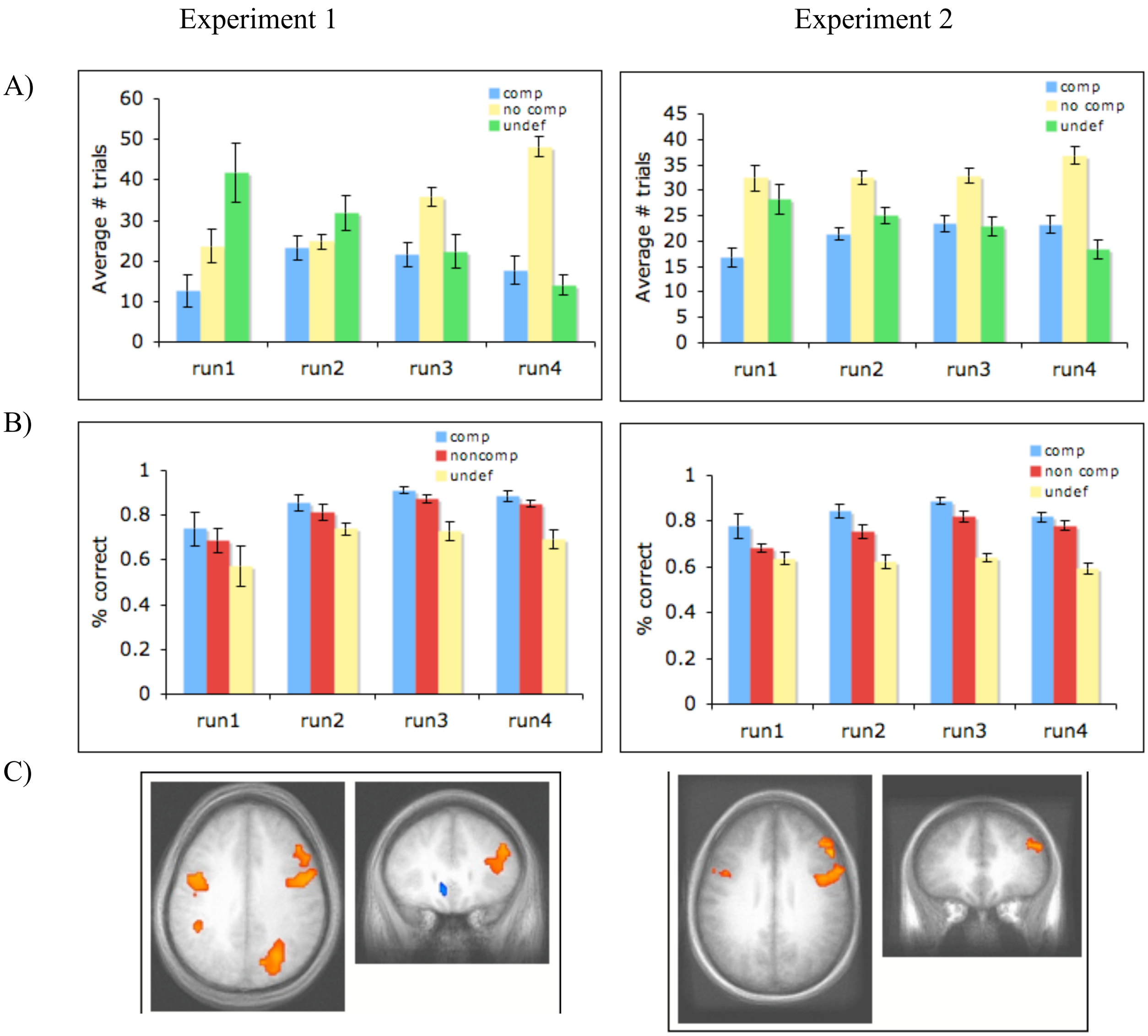
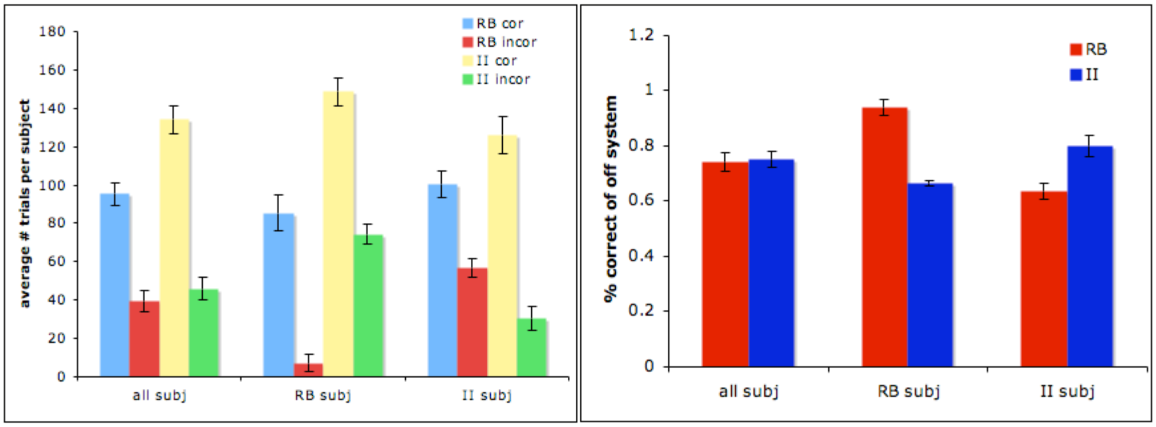


Figure 5.14: Competition-related activity

(A) Average number of Competition, Non-competition and Undefined trial types in Experiment 1 (left) and Experiment 2 (right). (B) Average accuracy of Competition, Non-competition and Undefined trial types in Experiment 1 (left) and Experiment 2 (right). (C) fMRI contrast of C vs. NC trial types in Experiment 1 (left) and Experiment 2 (right). The consistent regions of activity across studies occurs in the right DLPFC which we hypothesize corresponds to the operation of the decision module on these trials.

A)



B)

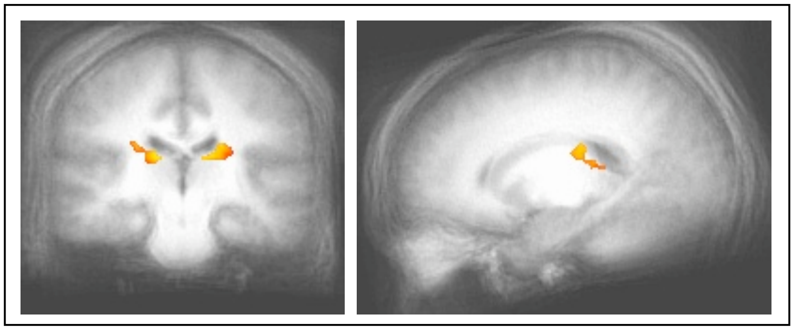


Figure 5.15: Off system activity in RB subjects

In the RB subject group in Experiment 2, trials were marked as correct or incorrect according to the predictions of the II system when PINNACLE predicted the RB system was active. A) The average number of trials in each subject group (left) and the accuracy of the 'on system' RB trials contrasted with the 'off system' II trials (right). C) The functional activity in the caudate ROI of RB subjects during II trials when the II system was not selected. The correct and incorrect designations are based on the II system's predictions, not the overt feedback given to the subject. These posterior caudate regions are consistent with the hypothesized 'II network', so the observation of activity here suggests that the II system is operating simultaneously with the RB system in these subjects.

CHAPTER 6 FIGURES

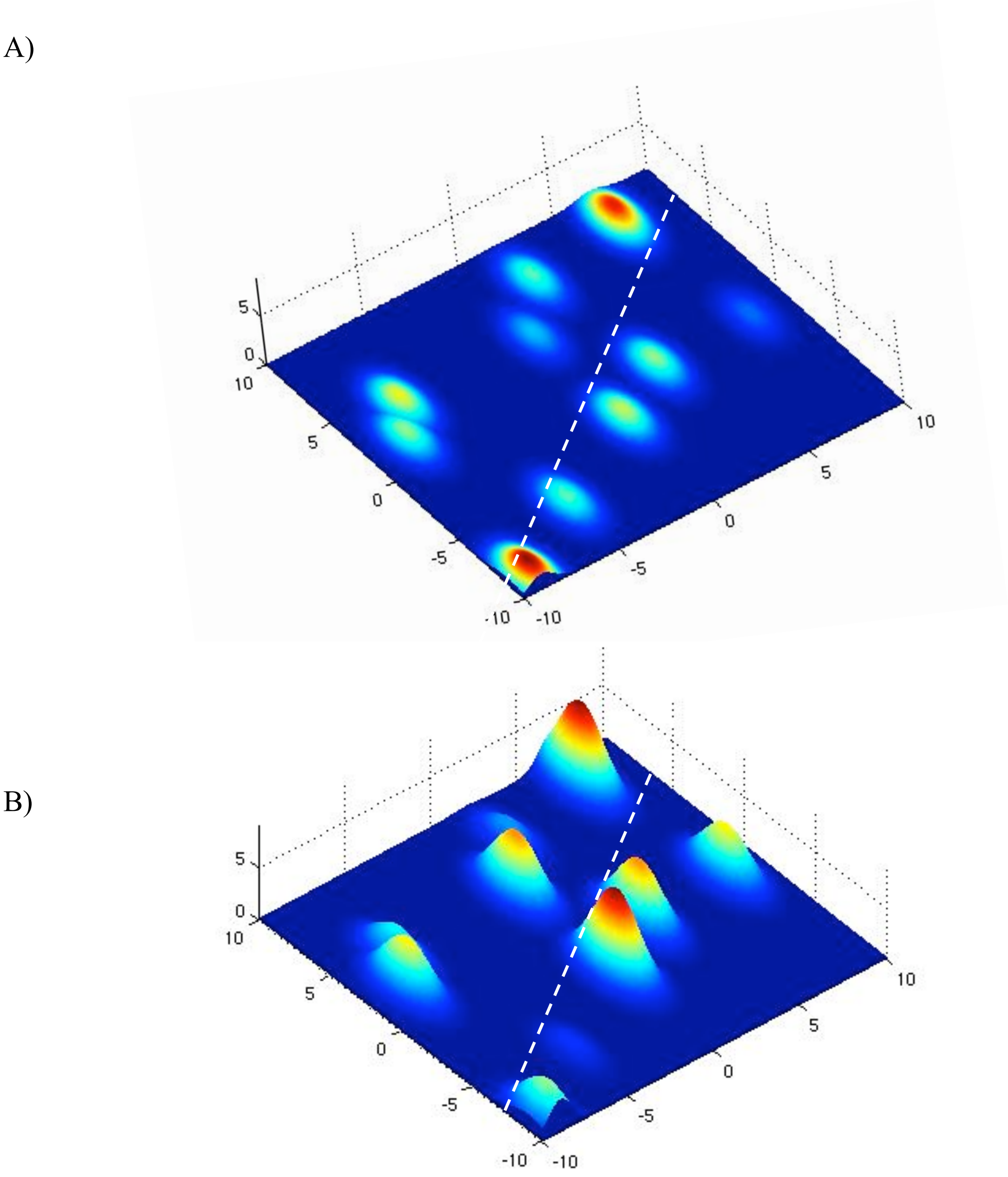


Figure 6.1: Simulation of II system clustering representation

(A) Hypothetical representation of a subject acquiring an II category structure by recruiting nodes within the stimulus space. Initially the clusters are small. (B) After some amount of experience the representation becomes stronger and more specific for each region that has been consistently endorsed.

REFERENCES

Anderson JR (1991) The adaptive nature of human categorization. Psychological Review 98:409-429.

Ashby FG (1992) Multidimensional Models of Perception and Cognition. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.

Ashby FG, Townsend JT (1986) Varieties of perceptual independence. Psychol Rev 93:154-179.

Ashby FG, Gott RE (1988) Decision rules in the perception and categorization of multidimensional stimuli. J Exp Psychol Learn Mem Cogn 14:33-53.

Ashby FG, Maddox WT (1990) Integrating information from separable psychological dimensions. J Exp Psychol Hum Percept Perform 16:598-612.

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

Ashby FG, Waldron EM (1999) On the nature of implicit categorization. Psychon Bull Rev 6:363-378.

Ashby FG, Ell SW (2001) The neurobiology of human category learning. Trends Cogn Sci 5:204-210.

Ashby FG, Valentin VV (2005) Multiple systems of perceptual category learning: theory and cognitive tests. New York: Elsevier.

Ashby FG, Ennis JM (2006) The Role of the Basal Ganglia in Category Learning. In: The Psychology of Learning and Motivation (Ross BH, ed), pp 1-36. New York: Elsevier.

Ashby FG, O'Brien JB (2007) The effects of positive versus negative feedback on information-integration category learning. Perception & Psychophysics 69:865-878.

Ashby FG, Queller S, Berretty PM (1999) On the dominance of unidimensional rules in unsupervised categorization. Percept Psychophys 61:1178-1199.

Ashby FG, Maddox WT, Bohil CJ (2002) Observational versus feedback training in rule-based and information-integration category learning. Mem Cognit 30:666-677.

Ashby FG, Ell SW, Waldron EM (2003a) Procedural learning in perceptual categorization. Mem Cognit 31:1114-1125.

Ashby FG, Ennis JM, Spiering BJ (2007) A neurobiological theory of automaticity in perceptual categorization. Psychol Rev 114:632-656.

Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM (1998) A neuropsychological theory of multiple systems in category learning. Psychol Rev 105:442-481.

Ashby FG, Noble S, Filoteo JV, Waldron EM, Ell SW (2003b) Category learning deficits in Parkinson's disease. Neuropsychology 17:115-124.

Brown RG, Marsden CD (1988) 'Subcortical dementia': the neuropsychological evidence. Neuroscience 25:363-387.

Cohen NJ, Poldrack RA, Eichenbaum H (1997) Memory for items and memory for relations in the procedural/declarative memory framework. Memory 5:131-178.

Collins DL, Neelin P, Peters TM, Evans AC (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18:192-205.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162-173.

Difiglia M, Pasik T, Pasik P (1978) A Golgi study of afferent fibers in the neostriatum of monkeys. Brain Res 152:341-347.

Ell SW, Marchant NL, Ivry RB (2006) Focal putamen lesions impair learning in rule-based, but not information-integration categorization tasks. Neuropsychologia 44:1737-1751.

Filoteo JV, Maddox WT, Davis JD (2001a) Quantitative modeling of category learning in amnesic patients. J Int Neuropsychol Soc 7:1-19.

Filoteo JV, Maddox WT, Davis JD (2001b) A possible role of the striatum in linear and nonlinear category learning: evidence from patients with Huntington's disease. Behav Neurosci 115:786-798.

Filoteo JV, Maddox WT, Salmon DP, Song DD (2005) Information-integration category learning in patients with striatal dysfunction. Neuropsychology 19:212-222.

Filoteo JV, Maddox WT, Ing AD, Song DD (in press) Characterizing Rule-Based Category Learning Deficits in Patients with Parkinson's Disease. Neuropsychologia.

Finch DM (1996) Neurophysiology of converging synaptic inputs from the rat prefrontal cortex, amygdala, midline thalamus, and hippocampal formation onto single neurons of the caudate/putamen and nucleus accumbens. Hippocampus 6:495-512.

Floresco SB, Grace AA (2003) Gating of hippocampal-evoked activity in prefrontal cortical neurons by inputs from the mediodorsal thalamus and ventral tegmental area. J Neurosci 23:3930-3943.

Foerde K, Knowlton BJ, Poldrack RA (2006) Modulation of competing memory systems by distraction. Proc Natl Acad Sci U S A 103:11778-11783.

Freund TF, Powell JF, Smith AD (1984) Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines. Neuroscience 13:1189-1215.

Gabbott P, Headlam A, Busby S (2002) Morphological evidence that CA1 hippocampal afferents monosynaptically innervate PV-containing neurons and NADPH-diaphorase reactive cells in the medial prefrontal cortex (Areas 25/32) of the rat. Brain Res 946:314-322.

Heindel WC, Salmon DP, Shults CW, Walicke PA, Butters N (1989) Neuropsychological evidence for multiple implicit memory systems: a comparison of Alzheimer's, Huntington's, and Parkinson's disease patients. J Neurosci 9:582-587.

Henson RN, Cansino S, Herron JE, Robb WG, Rugg MD (2003) A familiarity signal in human anterior medial temporal cortex? Hippocampus 13:301-304.

Hintzman DL (1986) "Schema abstraction" in a multiple-trace memory model. Psychological Review 93:411-428.

Hopkins RO, Myers CE, Shohamy D, Grossman S, Gluck M (2004) Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. Neuropsychologia 42:524-535.

Houk JC (2005) Agents of the Mind. Biol Cybern 92:427-437.

Houk JC, Bastianen C, Fansier D, Fishbach A, Fraser D, Reber PJ, Roy SA, Simo LS (2007) Action selection and refinement in subcortical loops through basal ganglia and cerebellum. Philos Trans R Soc Lond B 362:1573-1583.

Iaria G, Petrides M, Dagher A, Pike B, Bohbot VD (2003) Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. J Neurosci 23:5945-5952.

Insausti R, Juottonen K, Soininen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkanen A (1998) MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. AJNR Am J Neuroradiol 19:659-671.

Janowsky JS, Shimamura AP, Kritchevsky M, Squire LR (1989) Cognitive impairment following frontal lobe damage and its relevance to human amnesia. Behav Neurosci 103:548-560.

Jenkins IH, Brooks DJ, Nixon PD, Frackowiak RS, Passingham RE (1994) Motor sequence learning: a study with positron emission tomography. J Neurosci 14:3775-3790.

Joel D, Weiner I (2000) The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. Neuroscience 96:451-474.

Kelly RM, Strick PL (2004) Macro-architecture of basal ganglia loops with the cerebral cortex: use of rabies virus to reveal multisynaptic circuits. Prog Brain Res 143:449-459.

Kitchener EG, Squire LR (2000) Impaired verbal category learning in amnesia. Behav Neurosci 114:907-911.

Knowlton BJ, Squire LR (1993) The learning of categories: parallel brain systems for item memory and category knowledge. Science 262:1747-1749.

Knowlton BJ, Ramus SJ, Squire LR (1992) Intact artificial grammar learning in amnesia: Dissociation of category-level knowledge and explicit memory for specific instances. Psychol Sci 3:172-179.

Knowlton BJ, Squire LR, Gluck MA (1994) Probabilistic classification learning in amnesia. Learn Mem 1:106-120.

Knowlton BJ, Mangels JA, Squire LR (1996a) A neostriatal habit learning system in humans. Science 273:1399-1402.

Knowlton BJ, Squire LR, Paulsen JS (1996b) Dissociations within nondeclarative memory in Huntington's disease. Neuropsychology 10:538-548.

La Grutta V, Sabatino M (1988) Focal hippocampal epilepsy: effect of caudate stimulation. Experimental Neurology 99:38-49.

Law JR, Flanery MA, Wirth S, Yanike M, Smith AC, Frank LM, Suzuki WA, Brown EN, Stark CE (2005) Functional magnetic resonance imaging activity during the gradual acquisition and expression of paired-associate memory. J Neurosci 25:5720-5729.

Leng NR, Parkin AJ (1988) Double dissociation of frontal dysfunction in organic amnesia. Br J Clin Psychol 27 (Pt 4):359-362.

Lewis BL, O'Donnell P (2000) Ventral tegmental area afferents to the prefrontal cortex maintain membrane potential 'up' states in pyramidal neurons via D(1) dopamine receptors. Cereb Cortex 10:1168-1175.

Lombardi WJ, Andreason PJ, Sirocco KY, Rio DE, Gross RE, Umhau JC, Hommer DW (1999) Wisconsin Card Sorting Test performance following head injury: dorsolateral fronto-striatal circuit activity predicts perseveration. J Clin Exp Neuropsychol 21:2-16.

Love BC, Medin DL, Gureckis TM (2004) SUSTAIN: A Network Model of Category Learning. Psychological Review 111:309-332.

Maddox WT, Filoteo JV (2001) Striatal contributions to category learning: quantitative modeling of simple linear and complex nonlinear rule learning in patients with Parkinson's disease. J Int Neuropsychol Soc 7:710-727.

Maddox WT, Ashby FG (2004) Dissociating explicit and procedural-learning based systems of perceptual category learning. Behav Processes 66:309-332.

Maddox WT, Ing AD (2005) Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. J Exp Psychol Learn Mem Cogn 31:100-107.

Maddox WT, Molis MR, Diehl RL (2002) Generalizing a neuropsychological model of visual categorization to auditory categorization of vowels. Percept Psychophys 64:584-597.

Maddox WT, Ashby FG, Bohil CJ (2003) Delayed feedback effects on rule-based and information-integration category learning. J Exp Psychol Learn Mem Cogn 29:650-662.

Maddox WT, Bohil CJ, Ing AD (2004a) Evidence for a procedural-learning-based system in perceptual category learning. Psychon Bull Rev 11:945-952.

Maddox WT, Filoteo JV, Lauritzen JS (2007) Within-Category Discontinuity Interacts with Verbal Rule Complexity in Perceptual Category Learning. Journal of Experimental Psychology: Learning, Memory, and Cognition 33:197-218.

Maddox WT, Filoteo JV, Hejl KD, Ing AD (2004b) Category number impacts rule-based but not information-integration category learning: further evidence for dissociable category-learning systems. J Exp Psychol Learn Mem Cogn 30:227-245.

Maddox WT, Ashby FG, Ing AD, Pickering AD (2004c) Disrupting feedback processing interferes with rule-based but not information-integration category learning. Mem Cognit 32:582-591.

Maddox WT, Aparicio P, Marchant NL, Ivry RB (2005a) Rule-based category learning is impaired in patients with Parkinson's disease but not in patients with cerebellar disorders. J Cogn Neurosci 17:707-723.

Maddox WT, Filoteo JV, Lauritzen JS, Connally E, Hejl KD (2005b) Discontinuous categories affect information-integration but not rule-based category learning. J Exp Psychol Learn Mem Cogn 31:654-669.

Marr D, Poggio T (1976) Cooperative computation of stereo disparity. Science 194:283-287.

McDonald RJ, White NM (1993) A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. Behav Neurosci 107:3-22.

McDonald RJ, White NM (1994) Parallel information processing in the water maze: Evidence for independent memory systems involving dorsal striatum and hippocampus. Behavioral & Neural Biology 61:260-270.

McDonald RJ, White NM (1995) Hippocampal and non-hippocampal contributions to place learning. Behavioral Neuroscience 109:579-593.

Medin DL, Schaffer MM (1978) Context theory of classification learning. Psychological Review 5:207-238.

Meyer WJ, Offenbach SI (1962) Effectiveness of reward and punishment as a function of task complexity. Journal of Comparitive & Physiological Psychology 55:532-534.

Middleton FA, Strick PL (2000) Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Res Brain Res Rev 31:236-250.

Moody TD, Bookheimer SY, Vanek Z, Knowlton BJ (2004) An implicit learning task activates medial temporal lobe in patients with Parkinson's disease. Behav Neurosci 118:438-442.

Nomura EM, Maddox WT, Filoteo JV, Ing AD, Gitelman DR, Parrish TB, Mesulam MM, Reber PJ (2007) Neural correlates of rule-based and information-integration visual category learning. Cereb Cortex 17:37-43.

Nosofsky R, Zaki S (1999) Math modeling, neuropsychology, and category learning. Trends Cogn Sci 3:125-126.

Nosofsky RM (1986) Attention, similarity, and the identification-categorization relationship. J Exp Psychol Gen 115:39-61.

Olton DS, Samuelson RJ (1976) Rememberance of places passed: Spatial memory in rats. J Exp Psychol 2:97-115.

Packard MG (1999) Glutamate infused posttraining into the hippocampus or caudate-putamen differentially strengthens place and response learning. Proc Natl Acad Sci U S A 96:12881-12886.

Packard MG, McGaugh JL (1996) Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiol Learn Mem 65:65-72.

Packard MG, Teather LA (1999) Dissociation of multiple memory systems by post-training intracerebral injections of glutamate. Psychobiology 27:40-50.

Packard MG, Hirsh R, White NM (1989) Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: evidence for multiple memory systems. J Neurosci 9:1465-1472.

Poldrack RA, Gabrieli JD (2001) Characterizing the neural mechanisms of skill learning and repetition priming: evidence from mirror reading. Brain 124:67-82.

Poldrack RA, Rodriguez P (2004) How do memory systems interact? Evidence from human classification learning. Neurobiol Learn Mem 82:324-332.

Poldrack RA, Prabhakaran V, Seger CA, Gabrieli JD (1999) Striatal activation during acquisition of a cognitive skill. Neuropsychology 13:564-574.

Poldrack RA, Clark J, Pare-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, Gluck MA (2001) Interactive memory systems in the human brain. Nature 414:546-550.

Posner MI, Keele SW (1968) On the genesis of abstract ideas. J Exp Psychol 77:353-363.

Press W, Teukolsky S, Vetterling W, Flannery B (1992) Numerical Recipes in C, 2 Edition. Cambridge, MA: Cambridge University Press.

Rao SM, Bobholz JA, Hammeke TA, Rosen AC, Woodley SJ, Cunningham JM, Cox RW, Stein EA, Binder JR (1997) Functional MRI evidence for subcortical participation in conceptual reasoning skills. Neuroreport 8:1987-1993.

Rauch SL, Whalen PJ, Savage CR, Curran T, Kendrick A, Brown HD, Bush G, Breiter HC, Rosen BR (1997) Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. Hum Brain Mapp 5:124-132.

Reber PJ, Knowlton BJ, Squire LR (1996) Dissociable properties of memory systems: differences in the flexibility of declarative and nondeclarative knowledge. Behav Neurosci 110:861-871.

Reber PJ, Wong EC, Buxton RB (2002) Encoding activity in the medial temporal lobe examined with anatomically constrained fMRI analysis. Hippocampus 12:363-376.

Reber PJ, Gitelman DR, Parrish TB, Mesulam MM (2003) Dissociating explicit and implicit category knowledge with fMRI. J Cogn Neurosci 15:574-583.

Reed JM, Squire LR, Patalano AL, Smith EE, Jonides J (1999) Learning about categories that are defined by object-like stimuli despite impaired declarative memory. Behav Neurosci 113:411-419.

Reed SK (1972) Pattern recognition and categorization. Cognitive Psychology 3:382-407.

Rogers RD, Andrews TC, Grasby PM, Brooks DJ, Robbins TW (2000) Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. J Cogn Neurosci 12:142-162.

Schroeder JP, Wingard JC, Packard MG (2002) Post-Training Reversible Inactivation of Hippocampus Reveals Interference Between Memory Systems. Hippocampus 12:280-284.

Schultz W (1992) Activity of dopamine neurons in the behaving primate. Seminars in Neuroscience 4:129-138.

Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. J Neurol Neurosurg Psychiatry 20:11-21.

Seger CA, Cincotta CM (2005) The roles of the caudate nucleus in human classification learning. J Neurosci 25:2941-2951.

Seger CA, Cincotta CM (2006) Dynamics of frontal, striatal, and hippocampal systems during rule learning. Cereb Cortex 16:1546-1555.

Smiley JF, Levey AI, Ciliax BJ, Goldman-Rakic PS (1994) D1 dopamine receptor immunoreactivity in human and monkey cerebral cortex: predominant and extrasynaptic localization in dendritic spines. Proc Natl Acad Sci U S A 91:5720-5724.

Smith JD, Murray MJ, Minda JP (1997) Straight talk about linear separability. Memory and Cognition 23:659-680.

Sorensen KE, Witter MP (1983) Entorhinal efferents reach the caudate-putamen. Neuroscience Letters 35:259-264.

Squire LR (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol Rev 99:195-231.

Stark CE, Okada Y (2003) Making memories without trying: medial temporal lobe activity associated with incidental memory formation during recognition. The Journal of Neuroscience 23:6748-6753.

Sylvester CY, Wager TD, Lacey SC, Hernandez L, Nichols TE, Smith EE, Jonides J (2003) Switching attention and resolving interference: fMRI measures of executive functions. Neuropsychologia 41:357-370.

Tranel D, Damasio AR, Damasio H, Brandt JP (1994) Sensorimotor skill learning in amnesia: additional evidence for the neural basis of nondeclarative memory. Learn Mem 1:165-179.

Viskontas IV, Knowlton BJ, Steinmetz PN, Fried I (2006) Differences in mnemonic processing by neurons in the human hippocampus and parahippocampal regions. J Cogn Neurosci 18:1654-1662.

Voermans NC, Petersson KM, Daudey L, Weber B, Van Spaendonck KP, Kremer HP, Fernandez G (2004) Interaction between the human hippocampus and the caudate nucleus during route recognition. Neuron 43:427-435.

Waldron EM, Ashby FG (2001) The effects of concurrent task interference on category learning: evidence for multiple category learning systems. Psychon Bull Rev 8:168-176.

White NM, McDonald RJ (1993) Acquisition of a spatial conditioned place preference is impaired by amygdala lesions and improved by fornix lesions. Behav Brain Res 55:269-281.

White NM, McDonald RJ (2002) Multiple parallel memory systems in the brain of the rat. Neurobiol Learn Mem 77:125-184.

Wickens J (1990) Striatal dopamine in motor activation and reward-mediated learning: steps towards a unifying model. J Neural Transm Gen Sect 80:9-31.

Willingham DB (1998) A neuropsychological theory of motor skill learning. Psychol Rev 105:558-584.

Willingham DB, Salidas J, Gabrieli JDE (2002) Direct comparison of neural systems mediating conscious and unconscious skill learning. Journal of Neurophysiology 88:1451-1460.

Willingham DB, Wells LA, Farrell JM, Stemwedel ME (2000) Implicit motor sequence learning is represented in response locations. Mem Cognit 28:366-375.

Wilson C (1995) The contribution of cortical neurons to the firing pattern of striatal spiny neurons. Cambridge, MA: Bradford.

Yeterian EH, Pandya DN (1995) Corticostriatal connections of extrastriate visual areas in rhesus monkeys. J Comp Neurol 352:436-457.

Yeung N, Nystrom LE, Aronson JA, Cohen JD (2006) Between-task competition and cognitive control in task switching. J Neurosci 26:1429-1438.

Zeithamova D, Maddox WT (2006) Dual-task interference in perceptual category learning. Mem Cognit 34:387-398.

Zemel RS, Mozer MC (2001) Localist attractor networks. Neural Comput 13:1045-1064.

CURRICULUM VITAE

Emi M. Nomura

Northwestern University
Department of Psychology
2029 Sheridan Road
Evanston, IL 60201
Phone: (847) 467-5779
Fax: (847) 491-7589
e-nomura@northwestern.edu

RESEARCH INTERESTS

Cognitive neuroscience of category learning, functional MRI, computational modeling, neural system interactions

EDUCATION

2002 – 2008	Ph.D. in Neuroscience February 2008	Northwestern University; Chicago, IL Advisor: Paul J. Reber, PhD
1998 – 2002	B.A. (with honors) Major: Cognitive Science	Rice University; Houston, TX Advisor: Geoffrey Potts, PhD

PROFESSIONAL EXPERIENCE

2003 - 2008	**Cognitive Neuroscience of Memory Laboratory,** Northwestern University, Department of Psychology **Graduate Student**: laboratory of Dr. Paul Reber fMRI data collection and analysis with AFNI (Analysis of Functional Neuroimages) and bash software Computational modeling with python and MATLAB
2003 Winter	**Auditory Learning Laboratory**, Northwestern University, Department of Communication Sciences and Disorders **Sleep and Circadian Rhythms Laboratory**, Northwestern University, Department of Neurobiology and Physiology **Graduate student rotation**: between the laboratories of Dr. Beverley Wright and Dr. Phyllis Zee Examined the effect of sleep deprivation on an auditory discrimination learning task.
2002 Fall	**Attention and Motivation fMRI Laboratory,** Northwestern University, Department of Neurology **Graduate student rotation:** laboratory of Dr. Dana Small Used SPM99 fMRI data analysis software to look at shifts in spatial attention under rewarding and punishing conditions
2000 - 2002	**Motivation EEG Laboratory,** Rice University, Department of Psychology **Research assistant**: laboratory Dr. Geoffery Potts Applied EEG electrode nets, conducted experiments, analyzed EEG data

	Completed a Senior Honors Thesis project entitled Impulsivity Moderates the Brain's Response to Rewarding and Punishing Stimuli.
2001 Summer	**The University of Texas-Houston Summer Research Program,** Department of Neurobiology and Anatomy **Research Assistant**: laboratory of Dr. Jocelyne Bachevalier Assisted in the research of the ontogenetic development and decline of memory functions in non-human primates.
2000 Summer	**Sleep and Chronobiology Summer Behavioral Science Apprenticeship**, Brown University **Research Assistant**: laboratory of Dr. Mary Carskadon Studied the effect of sleep deprivation on adolescents through EEG recordings and melatonin samples. Completed a final research project on the biological aspect of insomnia and attended a professional meeting (APSS).
2000 Spring	**The Adolescent Motherhood Project,** Baylor College of Medicine **Research Assistant**: laboratory of Dr. Connie Wiemann Conducted a study on the prevalence of drug and alcohol abuse and depression among teenage mothers.

INVITED TALKS:

Functional Neuroimaging of Declarative and Non-declarative Memory Processes in Visual Category Learning. (2007) Northwestern University Mechanisms of Aging and Dementia Training Grant Day

fMRI of Declarative and Non-declarative Memory Processes in Visual Category Learning**.** (2006) Annual Northwestern University Institute for Neuroscience retreat

fMRI of visual category learning. (2006) Evanston Northwestern Hospital Radiology Department.

fMRI of Declarative and Non-declarative Memory Processes in Visual Category Learning. (2006) Northwestern University Cognitive Neurology and Alzheimer's Disease Center Day.

LIST OF PUBLICATIONS:

Published or in press

Nomura EM, Maddox WT, Filoteo JV, Ing AD, Gitelman DR, Parrish TB, Mesulam MM, Reber PJ. (2007). Neural correlates of rule-based and information integration categorization. Cerebral Cortex. 17(1): 37-43.

Nomura EM, Reber PJ. (2008) Neuroscience and Biobehavioral Reviews. Review of medial temporal lobe and caudate contributions to visual category learning. 32: 279-291.

Nomura EM, Maddox WT, Reber PJ. (in press) Cognitive Science Society. Cognitive model of visual category learning enhances fMRI data analysis.

Submitted or in preparation

Nomura EM, Reber PJ. (in prep) Neuroimaging and computational model of category learning combine to reveal new components neural networks supporting category learning.

Nomura EM, Reber PJ. (in prep) PINNACLE: A cognitive model of category learning.

CONFERENCE PRESENTATIONS:

Slide presentations

Nomura EM, Maddox WT, Reber PJ (2007). Cognitive model of visual category learning enhances fMRI data analysis. Cognitive Science Society Meeting.

Nomura EM, Maddox WT, Filoteo JV, Gitelman DR, Parrish TB, Mesulam MM, Reber PJ (2006). MTL and caudate contributions to visual category learning: combining fMRI and computational modeling. Slide presentation at the 36th Annual Meeting of the Society for Neuroscience.

Nomura EM. Cognitive Neuroscience of Category Learning Workshop (New York, NY, 2003). J.S. McDonnell Summer Student Fellow slide presentation titled "Neural Correlates of Rule-based and Information-Integration Categorization."

Poster presentations

Nomura EM, Maddox WT, Gitelman DR, Mesulam MM, Reber PJ (2007). Trial-by-trial computational model of rule-based and information-integration category learning enhances fMRI data analysis. Poster presentation at the 13th Annual Meeting of the Organization for Human Brain Mapping.

Nomura EM, Maddox WT, Gitelman DR, Parrish TB, Mesulam MM, Reber PJ (2007). Trial-by-trial computational model of category learning applied to fMRI data. Poster presentation at the 14th Annual Meeting of the Cognitive Neuroscience Society.

Nomura EM, Murray KD, Reber PJ. (2005). A Monte Carlo simulation method for setting statistical thresholds in fMRI data analysis. Poster presentation at the 35th Annual Meeting of the Society for Neuroscience.

Nomura EM, Maddox WT, Ing AD, Filoteo JV, Reber PJ (2004). Neural correlates of rule-based and information integration category learning: converging results from fMRI and computational modeling. Poster presentation at the 34th Annual Meeting of the Society for Neuroscience.

Nomura EM, Maddox WT, Filoteo JV, Gitelman DR, Parrish TB, Mesulam MM, Reber PJ (2004). Neural correlates of rule-based and information integration (II) categorization: An event-related fMRI study of category learning. Poster presentation at the 11th Annual Meeting of the Cognitive Neuroscience Society.

George MR, Potts GF, **Nomura EM**. (2003). Impulsivity and the Error-Related Negativity. Poster presentation at the ECNS, EEG and Clinical Neuroscience Society 5th Annual Conference

Small DM, **Nomura EM**, Simmons K, Parrish T, Gitelman D, and Mesulam MM (2003). Differential effects of monetary incentive on visual spatial expectancy and disengagement: The anterior vs the posterior cingulate cortex. Poster presentation at the 9th Annual Meeting of the Organization for Human Brain Mapping.

Small DM, **Nomura EM**, Simmons K, Parrish T, Gitelman D, and Mesulam MM (2003). Differential effects of monetary incentive on visual spatial expectancy and disengagement: The anterior vs the posterior cingulate cortex. NeuroImage. 19(2), Supplement 1, e991-e992.

Bachevalier J, **Nomura EM**. (2001). Orbital Frontal Lesions in Monkeys Impaired Object Recognition Memory when assessed with Problem-Solving Task but not with Preferential Looking Task. NIH funded UTH 2001 Summer Program Book of Abstracts.

HONORS & AWARDS

2007	Travel grant for Cognitive Science Society Meeting
2006	Northwestern University Conference Travel Grant
2005 - 2007	Mechanisms of Aging and Dementia Training Grant (T32 AG20506)
2004	Travel grant for Cognitive Neuroscience of Category Learning Meeting
2004	Northwestern University Conference Travel Grant
2003 - 2004	Neuroscience in the early-years training grant (T32 AG020418)
2002	Fellowship for Graduate study at Northwestern University
2002	Honors Project in Cognitive Science at Rice University
2001	Selected in an NIH funded program as a summer research intern
2000 – 2001	Selected for the National Dean's List
2001	President's Honor Roll
2000	President's Honor Roll
1999	President's Honor Roll

TEACHING EXPERIENCE:

2008 Spring	Guest lecturer for Aryeh Routtenberg, PhD in Introduction to Neuroscience
2007 Fall	T.A. to Aryeh Routtenberg, PhD in Introduction to Neuroscience
2005 Winter	T.A. to Satoro Suzuki, PhD. In Statistical Methods in Psychology

2002 - present	Science and Engineering Research and Teaching Synthesis tutor

SYNERGISTIC ACTIVITIES

2003 - 2005	South Chicago area High School Science Fair judge
2003 - 2006	Northwestern University Institute for Neuroscience social committee
2002 – 2005	Northwestern University Institute for Neuroscience recruiting committee

PROFESSIONAL ASSOCIATIONS

2007	Cognitive Science Society
2007	Organization for Human Brain Mapping *(Student Member)*
2001 – 2007	Society for Neuroscience *(Student Member)*
2003 – 2007	Cognitive Neuroscience Society *(Student Member)*
2000 – 2001	Sleep Research Society *(Student Member)*