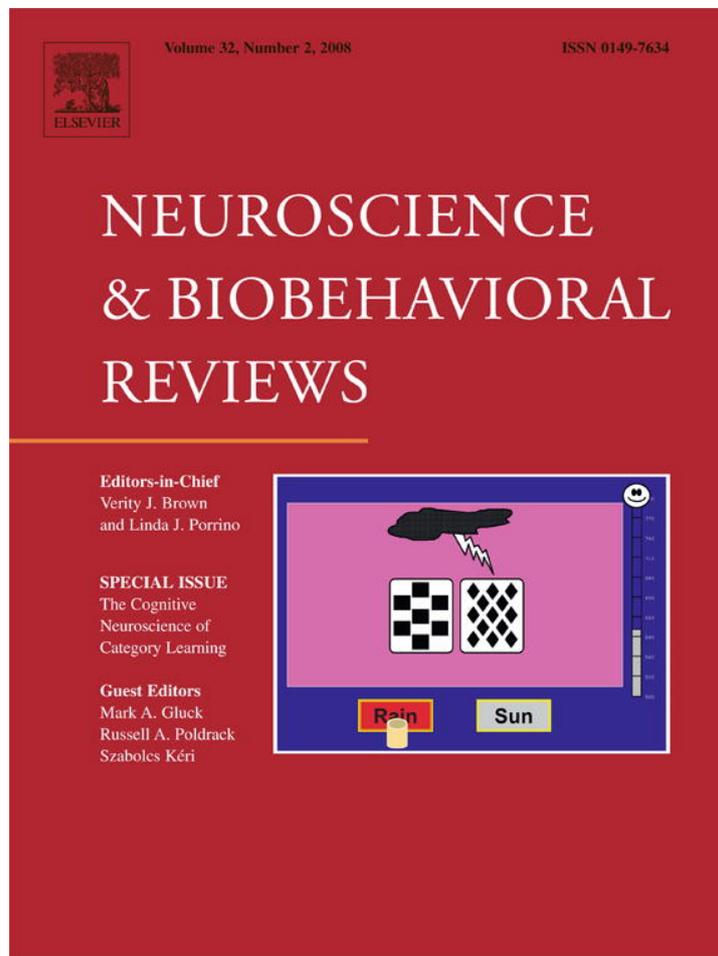


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER

Available online at www.sciencedirect.com

Neuroscience and Biobehavioral Reviews 32 (2008) 279–291

 NEUROSCIENCE AND
 BIOBEHAVIORAL
 REVIEWS

www.elsevier.com/locate/neubiorev

Review

A review of medial temporal lobe and caudate contributions to visual category learning

E.M. Nomura^{a,b}, P.J. Reber^{a,b,*}^a*Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60201, USA*^b*Institute for Neuroscience, Northwestern University, Chicago, IL 60611, USA*

Abstract

Here we review recent functional neuroimaging, neuropsychological and behavioral studies examining the role of the medial temporal lobe (MTL) and the caudate in learning visual categories either by verbalizable rules or without awareness. The MTL and caudate are found to play dissociable roles in different types of category learning with successful rule-based (RB) categorization depending selectively on the MTL and non-verbalizable information-integration (II) category learning depending on the posterior caudate. These studies utilize a combination of experimental cognitive psychology, mathematical modeling (Decision Bound Theory (DBT)) and cognitive computational modeling (the COVIS model of Ashby et al. [1998. A neuropsychological theory of multiple systems in category learning. *Psychological Review* 105, 442–481]) to enhance the understanding of data obtained via functional magnetic resonance imaging (fMRI). The combination of approaches is used to both test hypotheses of the cognitive model and also to incorporate hypotheses about the strategies used by participants to direct analysis of fMRI data. Examination of the roles of the MTL and caudate in visual category learning holds the promise of bridging between abstract cognitive models of behavior, systems neuroscience, neuropsychology, and the underlying neurophysiology of these brain regions.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Decision bound modeling; Categorization; Rule-based; Information-integration

Contents

1. Introduction	279
2. Brief review of RB and II behavioral experiments	282
3. Neuropsychological studies of RB/II category structures	284
4. Neuroimaging of category learning	285
5. Testing COVIS with functional magnetic resonance imaging (fMRI)	285
6. Mathematical modeling and fMRI data	287
7. The cognitive neuroscience of category learning	288
8. Future directions	290
Acknowledgements	290
References	290

1. Introduction

Categorization is a skill that allows us to respond similarly to distinct objects in the environment that share certain features. In visual categorization, novel stimuli are

*Corresponding author. Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60201, USA.
 Tel.: +1 847 467 1624; fax: +1 847 491 7859.

E-mail address: preber@northwestern.edu (P.J. Reber).

evaluated based on their perceptual features and treated as members of a category of related items (e.g., cats or dogs). Through experience with category members, new category representations can be formed that allow further identification of novel category members. The category learning process is a topic of broad and active investigation (Ashby and Maddox, 2005) for experimental cognitive psychology, computational cognitive models and cognitive neuroscience.

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 (1988). The basis of DBT is that people learn to assign motor responses to different regions in perceptual space. When presented with a to-be-categorized stimulus, subjects determine in what region the stimulus has fallen and produce the associated response. In this approach, learning the categories amounts to identifying the decision-boundary that separates the categories in the perceptual space. One consequence of this decision boundary is that those category members that are perceptually far from the boundary are categorized more easily and with higher confidence than those that are close to the boundary.

A number of reports supported DBT as an effective model of visual category learning (Ashby and Gott, 1988; Ashby and Maddox, 1990, 1992). Typically, the stimuli in these experiments vary on two dimensions. For example, in one task, subjects are asked to categorize rectangular stimuli that vary in either the length or the width (Fig. 1a). In another task, the stimuli are circles of different diameters that have an internal line that varies in orientation (Fig. 1b). The stimuli can also be perceptually more complex, such as sine wave gratings (Fig. 1c). All of

these examples can come from the same category structure, only differing in the stimulus dimensions. The two-dimensional perceptual space is partitioned into two (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.

A linear boundary that segments the perceptual space along one dimension (e.g., a horizontal or vertical boundary) creates two 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 two 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 Fig. 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 and may not be available to conscious awareness. As you can see in Fig. 2b, accurate categorization can only be achieved by incorporating both frequency and orientation information. Learning II category structures may rely on an implicit, procedural-learning-based system that gradually associates response labels with regions in stimulus space (Ashby and Waldron, 1999).

The Competition between Verbal and Implicit Systems model (COVIS model) proposed by Ashby et al. (1998) provides a specific hypothesis about the neural basis of RB and II categorization. In this model, two learning systems compete to provide the output response: an explicit, RB system dependent upon working memory and attention; and an implicit, II procedural learning system.

While the COVIS theory is based on the connections and computational properties of cortico-striatal 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 (1957; Squire, 1992) have established an important difference between conscious, declarative memory based on the medial temporal lobe (MTL) and a collection of heterogeneous non-declarative memory systems. Studies of non-declarative memory have shown the importance of the basal ganglia for some non-declarative memory tasks,

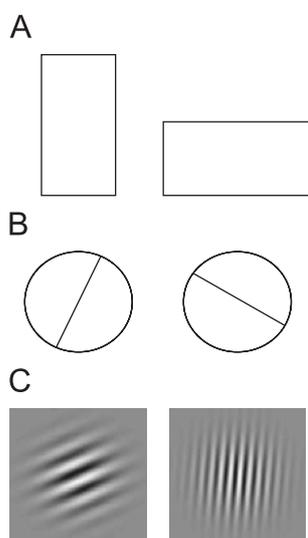


Fig. 1. (A) Rectangular stimuli that vary in width and length. (B) Circular stimuli that vary in diameter and line orientation. (C) Sine wave stimuli that vary in frequency and orientation.

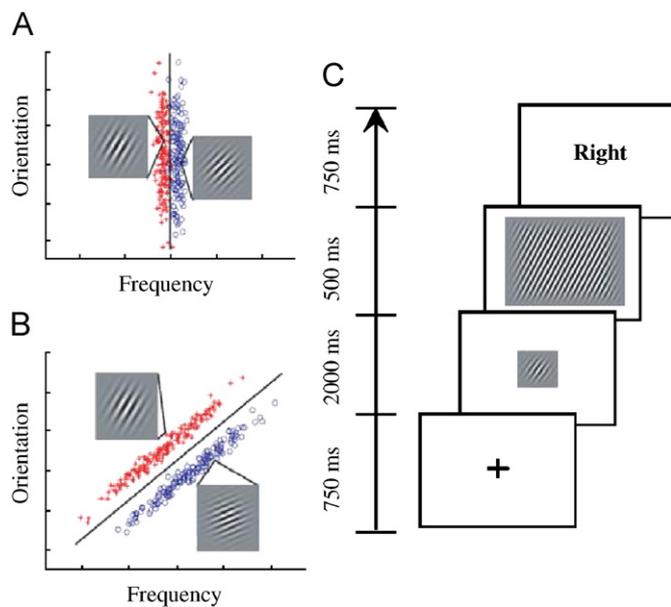


Fig. 2. RB (A) and II stimuli (B). Each point represents a distinct Gabor patch (sine-wave) stimulus defined by orientation (tilt) and frequency (thickness of lines). In both stimulus sets, there are two categories (red and blue points). RB categories are defined by a vertical boundary (only frequency 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 s 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.

including category learning tasks (Knowlton et al., 1996). Integrating these two conceptual frameworks is a fairly straightforward process.

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. Memory systems research emphasizes the critical role of the MTL in both acquisition and retrieval of conscious, declarative long-term knowledge that contributes to RB 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, in press). Because the MTL is hypothesized to be

exclusively involved in declarative memory supporting RB categorization, it is a crucial brain region for contrasting the neural correlates of RB and II category learning.

The II category 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 non-declarative 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). 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. There is some evidence from animal work that the glutamatergic 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).

The hypothesized organization of the neural circuits supporting RB and II category learning is shown in Fig. 3. This schematic is adapted from the COVIS theory of Ashby et al. (1998) but emphasizes the role of the MTL in supporting RB category learning as predicted by memory systems research. By this approach, RB categorization is thought to depend on learning within the PFC, MTL, head of the caudate, and the ACC. In contrast, II categorization

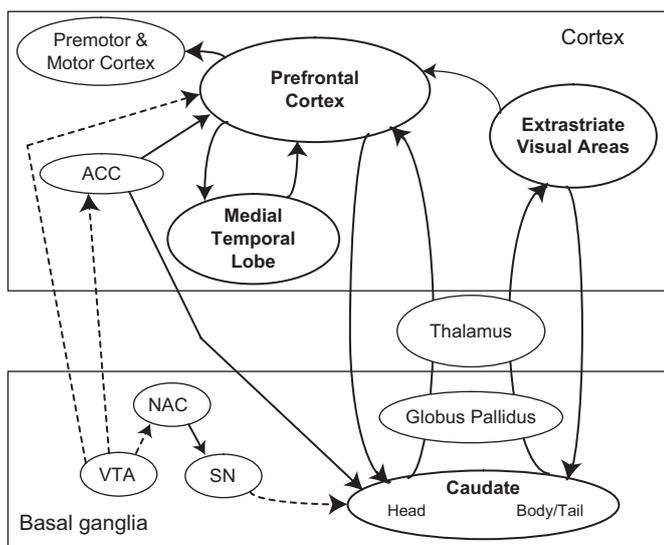


Fig. 3. Schematic of COVIS model adapted from Ashby et al. (1998) to incorporate reciprocal connections between the PFC and MTL (reflecting interactions found in research on declarative memory) as well as between extrastriate visual cortex and posterior regions of the caudate. RB category learning is 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. Anterior cingulate cortex (ACC); nucleus accumbens (NAC); substantia nigra (SN); and ventral tegmental area (VTA).

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.

While neurally inspired, the COVIS model makes several strong predictions about the different behavioral characteristics of RB and II category learning that have been tested empirically. The RB subsystem within COVIS depends on working memory (supported by PFC and ACC) for retrieval, comparison, evaluation and updating of the accuracy of the conscious rule in concert with the MTL for storage of the rule or exemplar. In contrast, the II subsystem is dependent on the specific timing of feedback, consistent category-response mappings and continuity of the category.

2. Brief review of RB and II behavioral experiments

A series of behavioral experiments have demonstrated dissociations between the RB and II category learning systems. Based on the predictions of the COVIS model, these experiments focus on four main manipulations: the nature and timing of the feedback, the presence of a concurrent task, 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 of COVIS is hypothesized to depend upon working memory and executive attention which allows 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 in order for dopamine to influence 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 et al. (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, people tend to rely on a 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. If feedback were delayed sufficiently from the presentation of the stimuli, II learning would be selectively impaired because of the inability to maintain information across the delay via working or long-term memory (which are exclusively involved in RB category learning). To test this hypothesis, 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 2.5s or more. This finding provides additional evidence for distinct neural systems driving RB and II category learning. Specifically, II learning depends on a procedural learning-based system that requires a tight temporal correlation between the motor response and the feedback for learning to take place, whereas RB learning relies on a hypothesis-testing system that is not time-dependent because of the ability to sustain information over longer times via working or long-term memory. This finding was replicated in an experiment that manipulated the feedback delay with four equally discriminable categories (Maddox and Ing, 2005). In the previous experiment, the II category boundary was oriented to require integration of two stimulus dimensions and the RB boundary only required attending to one stimulus dimension. With four categories, the number of relevant stimulus dimensions and thus attentional demands were equal between groups. Thus, we can conclude that the timing of feedback is an important quality of the II category learning

system irrespective of the number of categories being learned.

Because the RB system in COVIS depends on maintaining rules (and/or boundary reference information) in working memory during learning, the concurrent presence of a working memory task should preferentially impair RB but not II learning. This hypothesis was tested with a numerical analog of the Stroop task (Waldron and Ashby, 2001) and a sequential memory scanning task (Maddox and Ashby, 2004). Both studies found that RB learning was disrupted and II learning remained intact. A criticism of these studies is again that RB and II categorization depended on differing numbers of stimulus dimensions. A follow up study (Zeithamova and Maddox, 2006) examined the effect of varying the number of stimulus dimensions and found that even when RB learning was dependent upon two stimulus dimensions, the working memory task disrupted RB more than II learning. The transition between two and four categories should also involve an increase in working memory demands, and the results of this manipulation showed that RB learning was impacted more than II learning (Maddox et al., 2004b) with four categories.

The important role of feedback in II category learning suggests that it may rely on a type of procedural learning that shows similar feedback contingencies. Procedural learning is thought to reflect learning by doing, which in motor learning tasks relies upon the association between a particular motor pattern and the particular stimuli being learned (Willingham, 1998). In the serial reaction time (SRT) task, a well-known motor learning task, changing the location of the response keys interferes with learning but changing the sequence of finger movements does not (Willingham et al., 2000). One of the first studies to investigate the role of the particular motor action on category learning found that RB and II category learning were affected differently (Ashby et al., 2003a). In this experiment, subjects learned either RB or II category structures under three 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 two 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 et al.’s (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 RB category learning is mediated by an explicit hypothesis testing system that learns abstract category labels and II category learning is mediated by a procedural-learning-based system that associates category labels with response locations.

The fourth dissociation between RB and II category learning was observed by manipulating the distribution of stimuli within the categories (Maddox et al., 2005). Studies in both humans and animals indicate that procedural learning systems are dependent upon stimulus similarity and coherence (Cohen et al., 1997). Since II learning may be mediated by a procedural-learning-based system, stimuli that are grouped into discontinuous clusters should prove difficult in II not RB tasks. If RB category learning is dependent on an abstract verbal rule, it should be insensitive to discontinuities in the stimulus space. Rather, the rule could be applied regardless of the participant’s experience with the particular region in stimulus space. In this experiment, Maddox and colleagues compared RB and II group performance when the categories were either discontinuous or not. They found, as predicted, that assigning discontinuous clusters of stimuli to the same category label led to poor II category learning but had no effect on RB learning. In the same set of studies, RB learning transferred to novel parts of space whereas II learning did not, which strengthened the claim that II category learning is supported by a non-generalizable procedural learning system. A criticism of these results is that the RB categories contained only one relevant dimension, so the dissociation may be due to a difference in difficulty. In 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.

Across all of these studies, RB and II categorization were found to be dissociable as predicted by the COVIS model. However, while the behavioral dissociations demonstrate that the operating characteristics of RB and II category learning differ, they cannot directly address the hypotheses embedded in COVIS about the underlying neural systems that support RB and II category learning. Experimental techniques from cognitive neuroscience, i.e., neuropsychology and neuroimaging, are necessary to bridge from empirical cognitive psychology to neurobiology.

3. Neuropsychological studies of RB/II category structures

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 learning (Knowlton et al., 1992), and probabilistic classification (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 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 study using the same task, amnesics with focal bilateral hippocampal damage were impaired throughout the experiment (Hopkins et al., 2004), suggesting that the role of the hippocampus in this task may be complex. Across these studies, it appears that non-declarative memory may contribute to category learning. The COVIS model presents a specific hypothesis for a non-declarative memory mechanism for II category learning that depends on plasticity within cortico-striatal circuits, while RB category learning depends on declarative memory and the MTL.

The focus on the caudate in the COVIS model has meant that the majority of studies using RB and II tasks have focused on 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, there are also cognitive deficits. There is some disagreement in the literature as to the specific category learning deficit in PD patients. Learning in a perceptual categorization task that had simple line stimuli varying in horizontal and vertical length was impaired only with II, non-linear decision boundaries but was no different from controls with a RB, linear decision bound (Maddox and Filoteo, 2001). Similarly, Huntington's patients that also have damage to the striatum (Filoteo et al., 2001) show a selective impairment in II categorization.

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., 2005), but showed no impairment in II category learning. These conflicting results are difficult to resolve, but one explanation may be that the simple 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. In a follow-up study, 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.

The question of the RB category learning deficit that is present in some (Ashby et al., 2003b; Maddox et al., 2005) but absent in other studies (Maddox and Filoteo, 2001) can be resolved by grouping the results based on the number of irrelevant dimensions present in the task. The greater the number of irrelevant dimensions, the more selective attention that is required to focus on the relevant dimension. When examining those studies that observed an impairment in PD patients' ability to learn RB categories, we see that these were also the tasks with the greatest number of irrelevant stimulus dimensions. In a recent study, this issue was addressed by systematically manipulating the selective attention requirements during RB category learning. PD patients were found to be impaired at learning RB categories when selective attention demands were greatest (Filoteo et al., 2007). 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.

Data from patients with selective basal ganglia degeneration in simple category learning tasks is consistent with the hypothesis that the basal ganglia, the same structure that is implicated in procedural learning, is required for non-linear, II category learning. With increasing task complexity, the dissociation between RB and II category learning is 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.

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 verbalizable rules (Kitchener and Squire, 2000) although they can learn some complex categorization tasks implicitly (Reed et al., 1999). Amnesic patients thus appear to 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., 2001).

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.

4. Neuroimaging of 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 different neuroimaging studies have investigated the neural correlates of category learning in different types of tasks. In general, the data suggest that there exist multiple neural systems that can support category learning. 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.

One popular task is the probabilistic ‘weather prediction’ task first used by 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 are asked to determine 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.

Neuroimaging of the weather prediction task has consistently demonstrated the involvement of the basal ganglia (caudate nucleus) in 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, 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.

Another type of category learning paradigm is the dot pattern categorization task (Posner and Keele, 1968). In this task, participants are asked to observe a variety of dot patterns that have been distorted from an underlying prototype. In a neuroimaging study of the dot pattern 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.

From these disparate categorization paradigms, we can conclude that there are distinct neural structures supporting category learning depending on the particular type of learning required. It would seem that there is a declarative, RB neural network that can support explicit learning of associations or memorization of categorical patterns. There also exist at least two 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.

5. Testing COVIS with functional magnetic resonance imaging (fMRI)

Previous neuroimaging studies have demonstrated the involvement of multiple neural systems in RB and II category learning. While these studies identified some of the areas of activity predicted by the COVIS hypothesis, a clean dissociation between the RB and II systems has 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 T 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 (Fig. 2a), but in the II group, frequency and orientation information needed to be combined in order to determine category membership (Fig. 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.

The RB and II groups were scanned while categorizing the stimuli over four 80-trial runs. Feedback occurred immediately after each trial, enabling the participants to learn the categories while fMRI data were collected. The trial structure can be seen in Fig. 2c. Behaviorally, performance in both groups was above chance in all runs, average accuracy did not differ between groups ($F(1,24) = 1.25$, n.s.), and the groups showed similar learning curves. In the imaging data, for each group, activity evoked by correct categorization was compared to that evoked by incorrect categorization to identify brain areas associated with successful categorization. Monte Carlo simulation with matched noise data identified a reliability threshold of $t > 4.5$ ($p < .01$ uncorrected) for all voxels in clusters of at least 300 mm^3 that provides a brain-wide alpha rate of less than .05 false positives. RB categorization showed greater differential activation in the hippocampus (Fig. 4a) and additional areas of success-correlated activity in the ACC and medial frontal gyrus. For the II group there was greater differential activation in the head and tail of the caudate (Fig. 4b) for correct trials compared with incorrect trials.

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 MTL and the caudate. For each participant, ROIs were drawn following anatomical boundaries that are visible on

structural MRI and aligned using the ROI alignment (ROI-AL) method described elsewhere (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 whether there was significantly different activity associated with successful categorization in the RB and II groups in these regions. A double subtraction of correct versus incorrect trials and RB versus II groups revealed significant clusters of differential activity within the ROIs. Specifically, successful RB categorization evoked greater differential activation in the left anterior hippocampus than successful II categorization (Fig. 4c) and successful II categorization evoked greater differential activation in the right posterior caudate than successful RB categorization (Fig. 4d).

It is notable that the peri-stimulus time course (Fig. 4c, right) of evoked activity in the MTL cluster reflects a downward deflection for all trials. Unlike previous studies in which 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 represents 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 a few trials.

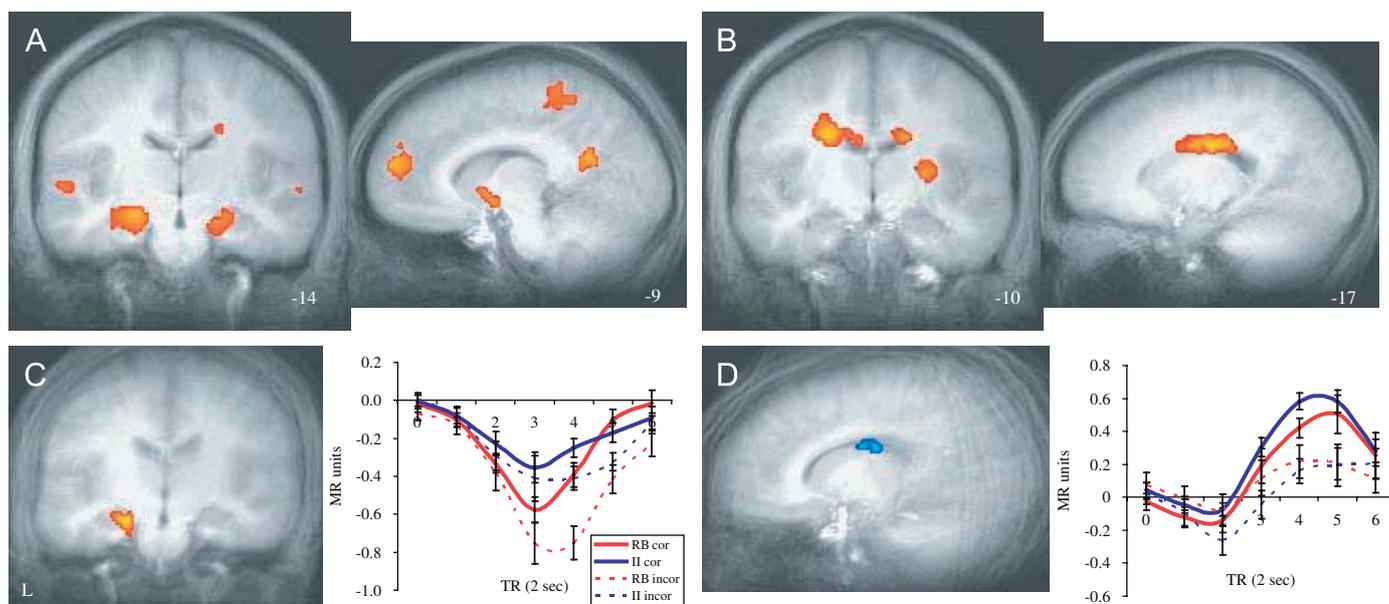


Fig. 4. (A) A portion of the successful RB categorization network featuring the MTL ($t(11) > 4.5$, cluster $> 300 \text{ mm}^3$). Sagittal view (right) shows increased activity in the anterior and posterior cingulate but not the caudate. (B) A portion of the successful II categorization network featuring the body of the caudate. ($t(12) > 4.5$, cluster $> 300 \text{ mm}^3$). (C) The RB successful categorization difference was reliably larger than the II successful categorization difference in the left MTL ($t(24) > 2.0$, cluster $> 700 \text{ mm}^3$); via ROI-AL analysis, Stark and Okada, 2003). Peri-stimulus activity time courses shown in the right panel. (D) The II successful categorization difference was reliably larger than the RB successful categorization difference in the right body of the caudate ($t(24) > 2.0$, cluster $> 600 \text{ mm}^3$); via ROI-AL analysis of the caudate). Peri-stimulus activity time courses shown in the right panel.

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 caudate body than in the RB group (Fig. 4d). The peri-stimulus time course (Fig. 4d, right) 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. A computational model based on the ideas put forth in COVIS can help address this issue.

6. Mathematical modeling and fMRI data

Modeling provides a more precise method for identifying strategy-use than the typical accuracy-based analyses. In the fMRI data discussed previously, all analyses were based on a subtraction of correct minus incorrect trials to isolate areas associated with successful category learning. While this is effective for a large group analysis, with computational modeling we can identify blocks of trials in individual subject data that demonstrate best RB or II strategy use. This grouping then provides us with a clearer image of activity associated specifically with RB or II category learning and may result in the visualization of a greater number of components of each of these systems. For example, modeling may identify areas such as the PFC and ACC in RB category learning and extrastriate visual cortex in II category learning that were not seen in the accuracy-based results. This approach is also important because participants in an II task may in fact be using an RB strategy or vice versa. Using the model-based analyses to isolate participants using different strategies provides a more rigorous window into potential brain activation and performance differences.

Here we have instantiated RB and II decision-bound models that are derived from general recognition theory (GRT; Ashby and Townsend 1986), which is a multivariate generalization of signal-detection theory. It is assumed that the percept can be represented as a point in a multi-dimensional 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 data for each of the four 80-trial blocks for each participant. 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) and each block of trials was fit to identify the placement of this boundary and perceptual noise parameter that best accounted for the observed data. The II model assumed a decision boundary with slope equal to 1.0 (i.e. a diagonal line reflecting integration of both dimensions) and the best fitting intercept and perceptual noise parameter combination was found for each block. Thus the models both had exactly two free parameters to allow for direct comparison of fit. 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 the best examples of the use of each strategy. We used this classification to organize the fMRI data collected previously and found that the model-identified period of best

strategy expression successfully identified additional elements of the neural networks for RB and II category learning. Fig. 5 contrasts the top 15 subject runs that were best fit by the RB model with the top 15 best fit by the II model for correct trials only (in some cases, subjects contributed more than one run to the analysis). Activity during correct trials was compared for RB and II performance because the well-fit runs tended to be those runs where there was a high degree of accuracy and few incorrect trials. During the runs where RB behavior was being most clearly expressed, increased activity was observed in the right PFC compared with the best II runs. In a right occipital cortical area, greater activity was observed during the examples of II strategy use compared with the best RB runs. These results are consistent with the hypothesis that the right PFC is involved when one is correctly utilizing an RB strategy. The increased activity in right occipital cortex during clear II strategy use suggests that this region may be the cortical element of the cortico-striatal connections hypothesized to represent category knowledge acquired via the II category learning system.

Although COVIS makes strong predictions regarding the cortical contributions to RB and II category learning, differential activity in these regions was not reliably observed when participants were simply grouped by condition without respect to their ability to clearly express the appropriate strategy. The successful application of the model-fitting technique in order to isolate the most effective applications of RB and II strategies suggests that modeling techniques in conjunction with fMRI analysis can be used to identify the detailed brain networks supporting cognitive processes.

7. The cognitive neuroscience of category learning

The COVIS model proposes a neurobiologically inspired instantiation of mechanisms for accomplishing DBT category learning with specific predictions about the neural systems supporting two separate types of category learning (RB and II). 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 (Nomura et al., 2007) shows that the roles of the MTL and the caudate can be doubly dissociated for RB and II category learning. These findings together with behavioral and neuropsychological evidence support the principle claims of COVIS that these two types of category learning are supported by separate neural systems with different operating characteristics.

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 a mathematical model that characterizes behavior such as DBT. As seen above, DBT provides a mechanism for examining the detailed behavioral data produced by a participant and conjecturing

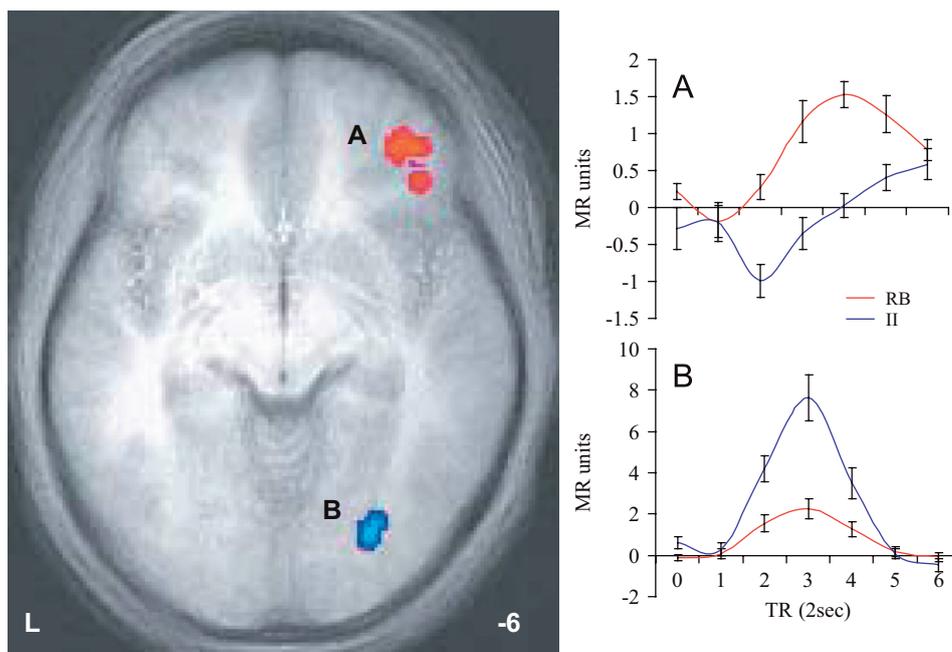


Fig. 5. Correct trial activity for model-identified best strategy use and the corresponding peri-stimulus time courses of activity within significant clusters: (A) best RB fit runs evoked greater activity than best II fit runs in right PFC. (B) Best II fit runs evoked greater activity than best RB fit runs in right occipital cortex. For this analysis, the statistical threshold was $t(30) > 3.5$, in a contiguous cluster of greater than 300 mm^3 .

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 tests whether the predictions of this type of cognitive model has a consistent neural correlate. The specific neural correlates thus identified provide additional tests 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 interesting features of the observed peri-stimulus time courses obtained with fMRI during category learning hold out the promise of further testing specific predictions of COVIS (or increasingly specific and detailed versions of COVIS). The de-activation in the MTL may indicate the type(s) of declarative memory function that support RB category learning. The evoked activity for all trials in the posterior caudate may provide some insight into the nature of the competition between the RB and II category learning systems. This activity appears to be evoked on every trial for both types of category learning, although the magnitude of the activity is only correlated with correct category judgments for the II task. The commonalities in the overall activity patterns in the MTL and caudate suggest that the competition between the two systems may be resolved via a higher-order decision process (e.g., in the PFC). This type of observation may serve to motivate additional neuroimaging studies aimed at testing increasingly specific hypotheses about the specific neural basis of the COVIS model.

In addition to using fMRI to test the COVIS model, cognitive neuroscience data may provide important constraints on testing which models of category learning best account for human behavior. Most cognitive models are not described at a level that incorporates an underlying neural basis. The development of the cognitive neuroscience of categorization will push the effective cognitive models to address neuroscientific data and constraints.

The DBT model of categorization is a relatively simple method for describing boundaries in perceptual space to represent visual categories. It has typically been used in paradigms in which the experimental stimuli can be described in a small number of dimensions (usually two). Other approaches to learning and representing categories in high-dimensional representational spaces include hierarchical clustering models such as SUSTAIN (Love et al., 2004) or Anderson's rational model of categorization (Anderson, 1991) and exemplar models (Nosofsky, 1985) such as ALCOVE (Kruschke, 1992). These models have typically been compared and contrasted based on quality of fit to behavioral performance of participants acquiring new categories in experimental situations. Adding an entirely new source of data, such as functional neuroima-

ging, should provide additional constraints on model-fitting that will better enable comparison between cognitive computational models.

Before functional neuroimaging can be applied to the question of assessing the merits of different cognitive models, hypotheses need to be developed about the relationship of the model structure to brain function. These sophisticated cognitive models contain strong hypotheses about representational structures but do not typically indicate how these structures are represented in neural systems. It is not immediately clear how SUSTAIN or the rational categorization model proposes to instantiate cluster representations and how these are acquired. However, the relationship of the DBT model to the COVIS theory can provide a framework for extending these models to the brain.

The neuroimaging results presented here 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 model, that model does not make specific computational predictions about the representations inherent in the RB and II subsystems. One possibility is to consider the II learning system as effecting a clustering learning algorithm that captures some important elements of those models. The Striatal Pattern Classifier (SPC) model (Ashby and Waldron, 1999) is an example of a model in which regions of perceptual space are labeled by II category learning. This process is hypothesized to be accomplished by cortico-striatal circuits involving the posterior regions of the caudate. Extending this approach to more abstract categories that may not have easily describable perceptual spaces would enable this type of mechanism to accomplish important elements of the cluster-based category learning models.

For a cognitive model such as SUSTAIN, a cluster-learning function based on the caudate would have to be proposed together with additional neural mechanisms for maintenance of goals, feature weighting and cluster hierarchies. However, the current results at least point to a direction for the development of a neural model of complex cognitive models of categorization.

Exemplar theory is typically presented as implying that long-term declarative memory for specific stimuli form the basis of category judgments. This view is difficult to reconcile with the observation of intact category learning in amnesic patients (Knowlton and Squire, 1993; Knowlton et al., 1994) but a mathematical model that attempts to do this has been described (Nosofsky and Zaki, 1998). If a hypothesis for how the model proposed by Nosofsky and Zaki 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 (e.g., Reber et al., 2003).

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

push towards unification of 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.

8. Future directions

Functional neuroimaging acts as an effective bridge to provide data that connects sophisticated cognitive models with the operation of the underlying neural structures. As the relationship between the details of the BOLD signal and the cell firing patterns that create the metabolic demands becomes better understood, it will be possible to learn more about the operations of the neural systems from the characteristics of the peri-stimulus time courses. Toward this end, we are developing a computational model that attempts 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 current decision structure that adjudicates competition between the models. As the computational models become more detailed and specific, these models provide an opportunity to test the psychological reality of the model by identifying if specific representations and parameters have real neural correlates. In addition, these models can enhance the analysis of functional neuroimaging data by identifying trials or blocks of trials that are particularly clear examples of specific strategies.

The cognitive neuroscience of category learning holds tremendous promise as a domain for interdisciplinary integration of a wide range of data and ideas from neurophysiology to systems neuroscience to cognitive and computational models of human behavior. Identifying the roles and cognitive functions of the MTL and the caudate in category learning provides a singular opportunity for bringing knowledge of those systems from neuropsychology and neurobiology to further inform theories of category learning.

Acknowledgements

The authors would like to acknowledge the funding sources for this research: NIMH R01-MH58748 (PJR), Mechanisms of Aging and Dementia Training Grant T32-AG020418 (EMN).

References

- Anderson, J.R., 1991. The adaptive nature of human categorization. *Psychological Review* 98, 409–429.
- Ashby, F.G., 1992. *Multidimensional Models of Perception and Cognition*. Lawrence Erlbaum Associates Publishers, Hillsdale, NJ.
- Ashby, F.G., Gott, R.E., 1988. Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning Memory and Cognition* 14, 33–53.
- Ashby, F.G., Maddox, W.T., 1990. Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance* 16, 598–612.
- Ashby, F.G., Maddox, W.T., 1992. Complex decision rules in categorization: contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance* 18, 50–71.
- Ashby, F.G., Maddox, W.T., 2005. Human category learning. *Annual Review of Psychology* 56, 149–178.
- Ashby, F.G., Townsend, J.T., 1986. Varieties of perceptual independence. *Psychological Review* 93 (2), 154–179.
- Ashby, F.G., Valentin, V.V., in press. *Multiple Systems of Perceptual Category Learning: Theory and Cognitive Tests*. Elsevier, New York.
- Ashby, F.G., Waldron, E.M., 1999. On the nature of implicit categorization. *Psychonomic Bulletin & Review* 6, 363–378.
- Ashby, F.G., Alfonso-Reese, L.A., Turken, A.U., Waldron, E.M., 1998. A neuropsychological theory of multiple systems in category learning. *Psychological Review* 105, 442–481.
- Ashby, F.G., Queller, S., Berretty, P.M., 1999. On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics* 61, 1178–1199.
- Ashby, F.G., Ell, S.W., Waldron, E.M., 2003a. Procedural learning in perceptual categorization. *Memory & Cognition* 31, 1114–1125.
- Ashby, F.G., Noble, S., Filoteo, J.V., Waldron, E.M., Ell, S.W., 2003b. Category learning deficits in Parkinson's disease. *Neuropsychology* 17, 115–124.
- Brown, R.G., Marsden, C.D., 1988. 'Subcortical dementia': the neuropsychological evidence. *Neuroscience* 25, 363–387.
- Cohen, N.J., Poldrack, R.A., Eichenbaum, H., 1997. Memory for items and memory for relations in the procedural/declarative memory framework. *Memory* 5, 131–178.
- Difiglia, M., Pasik, T., Pasik, P., 1978. A Golgi study of afferent fibers in the neostriatum of monkeys. *Brain Research* 152, 341–347.
- Filoteo, J.V., Maddox, W.T., Davis, J.D., 2001. Quantitative modeling of category learning in amnesic patients. *Journal of the International Neuropsychological Society* 7, 1–19.
- Filoteo, J.V., Maddox, W.T., Salmon, D.P., Song, D.D., 2005. Information-integration category learning in patients with striatal dysfunction. *Neuropsychology* 19, 212–222.
- Filoteo, J.V., Maddox, W.T., Ing, A.D., Song, D.D., 2007. Characterizing rule-based category learning deficits in patients with Parkinson's disease. *Neuropsychologia* 21, 183–192.
- Freund, T.F., Powell, J.F., Smith, A.D., 1984. Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines. *Neuroscience* 13, 1189–1215.
- Henson, R.N., Cansino, S., Herron, J.E., Robb, W.G., Rugg, M.D., 2003. A familiarity signal in human anterior medial temporal cortex? *Hippocampus* 13, 301–304.
- Hopkins, R.O., Myers, C.E., Shohamy, D., Grossman, S., Gluck, M., 2004. Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia* 42, 524–535.
- Janowsky, J.S., Shimamura, A.P., Kritchevsky, M., Squire, L.R., 1989. Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral Neuroscience* 103, 548–560.
- Kitchener, E.G., Squire, L.R., 2000. Impaired verbal category learning in amnesia. *Behavioral Neuroscience* 114, 907–911.
- Knowlton, B.J., Squire, L.R., 1993. The learning of categories: parallel brain systems for item memory and category knowledge. *Science* 262, 1747–1749.
- Knowlton, B.J., Ramus, S.J., Squire, L.R., 1992. Intact artificial grammar learning in amnesia: dissociation of category-level knowledge and explicit memory for specific instances. *Psychological Science* 3, 172–179.
- Knowlton, B.J., Squire, L.R., Gluck, M.A., 1994. Probabilistic classification learning in amnesia. *Learning Memory* 1, 106–120.

- Knowlton, B.J., Mangels, J.A., Squire, L.R., 1996. A neostriatal habit learning system in humans. *Science* 273, 1399–1402.
- Kruschke, J.K., 1992. ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review* 99, 22–44.
- Leng, N.R., Parkin, A.J., 1988. Double dissociation of frontal dysfunction in organic amnesia. *British Journal of Clinical Psychology* 27 (Part 4), 359–362.
- Lombardi, W.J., Andreason, P.J., Sirocco, K.Y., Rio, D.E., Gross, R.E., Umhau, J.C., Hommer, D.W., 1999. Wisconsin Card Sorting Test performance following head injury: dorsolateral fronto-striatal circuit activity predicts perseveration. *Journal of Clinical and Experimental Neuropsychology* 21, 2–16.
- Love, B.C., Medin, D.L., Gureckis, T.M., 2004. SUSTAIN: a network model of category learning. *Psychological Review* 111, 309–332.
- Maddox, W.T., Ashby, F.G., 2004. Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavior Processes* 66, 309–332.
- Maddox, W.T., Filoteo, J.V., 2001. Striatal contributions to category learning: quantitative modeling of simple linear and complex nonlinear rule learning in patients with Parkinson's disease. *Journal of the International Neuropsychological Society* 7, 710–727.
- Maddox, W.T., Ing, A.D., 2005. Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *Journal of Experimental Psychology: Learning Memory and Cognition* 31, 100–107.
- Maddox, W.T., Ashby, F.G., Bohil, C.J., 2003. Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning Memory and Cognition* 29, 650–662.
- Maddox, W.T., Bohil, C.J., Ing, A.D., 2004a. Evidence for a procedural-learning-based system in perceptual category learning. *Psychonomic Bulletin & Review* 11, 945–952.
- Maddox, W.T., Filoteo, J.V., Hejl, K.D., Ing, A.D., 2004b. Category number impacts rule-based but not information-integration category learning: further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning Memory and Cognition* 30, 227–245.
- Maddox, W.T., Aparicio, P., Marchant, N.L., Ivry, R.B., 2005. Rule-based category learning is impaired in patients with Parkinson's disease but not in patients with cerebellar disorders. *Journal of Cognitive Neuroscience* 17, 707–723.
- Maddox, W.T., Filoteo, J.V., Lauritzen, J.S., 2007. Within-category discontinuity interacts with verbal rule complexity in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33, 197–218.
- Middleton, F.A., Strick, P.L., 2000. Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Research. Brain Research Reviews* 31, 236–250.
- Moody, T.D., Bookheimer, S.Y., Vanek, Z., Knowlton, B.J., 2004. An implicit learning task activates medial temporal lobe in patients with Parkinson's disease. *Behavioral Neuroscience* 118, 438–442.
- Nomura, E.M., Maddox, W.T., Filoteo, J.V., Ing, A.D., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., Reber, P.J., 2007. Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex* 17, 37–43.
- Nosofsky, R., 1985. Overall similarity and the identification of separable-dimension stimuli: a choice model analysis. *Perception & Psychophysics* 38, 415–432.
- Nosofsky, R., Zaki, S., 1998. Dissociations between categorization and recognition in amnesic and normal individuals: an exemplar-based interpretation. *Psychological Science* 9, 247–255.
- Poldrack, R.A., Prabhakaran, V., Seger, C.A., Gabrieli, J.D., 1999. Striatal activation during acquisition of a cognitive skill. *Neuropsychology* 13, 564–574.
- Poldrack, R.A., Clark, J., Pare-Blagoev, E.J., Shohamy, D., Creso Moyano, J., Myers, C., Gluck, M.A., 2001. Interactive memory systems in the human brain. *Nature* 414, 546–550.
- Posner, M.I., Keele, S.W., 1968. On the genesis of abstract ideas. *Journal of Experimental Psychology* 77, 353–363.
- Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1992. *Numerical Recipes in C*, second ed. Cambridge University Press, Cambridge, MA.
- Rao, S.M., Bobholz, J.A., Hammeke, T.A., Rosen, A.C., Woodley, S.J., Cunningham, J.M., Cox, R.W., Stein, E.A., Binder, J.R., 1997. Functional MRI evidence for subcortical participation in conceptual reasoning skills. *Neuroreport* 8, 1987–1993.
- Rauch, S.L., Whalen, P.J., Savage, C.R., Curran, T., Kendrick, A., Brown, H.D., Bush, G., Breiter, H.C., Rosen, B.R., 1997. Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. *Human Brain Mapping* 5, 124–132.
- Reber, P.J., Knowlton, B.J., Squire, L.R., 1996. Dissociable properties of memory systems: differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral Neuroscience* 110, 861–871.
- Reber, P.J., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., 2003. Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience* 15, 574–583.
- Reed, J.M., Squire, L.R., Patalano, A.L., Smith, E.E., Jonides, J., 1999. Learning about categories that are defined by object-like stimuli despite impaired declarative memory. *Behavioral Neuroscience* 113, 411–419.
- Rogers, R.D., Andrews, T.C., Grasby, P.M., Brooks, D.J., Robbins, T.W., 2000. Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. *Journal of Cognitive Neuroscience* 12, 142–162.
- Schultz, W., 1992. Activity of dopamine neurons in the behaving primate. *Seminars in Neuroscience* 4, 129–138.
- Scoville, W.B., Milner, B., 1957. Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology Neurosurgery Psychiatry* 20, 11–21.
- Seger, C.A., Cincotta, C.M., 2005. The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience* 25, 2941–2951.
- Seger, C.A., Cincotta, C.M., 2006. Dynamics of frontal, striatal and hippocampal systems during rule learning. *Cerebral Cortex* 16 (11), 1546–1555.
- Smiley, J.F., Levey, A.I., Ciliax, B.J., Goldman-Rakic, P.S., 1994. D1 dopamine receptor immunoreactivity in human and monkey cerebral cortex: predominant and extrasynaptic localization in dendritic spines. *Proceedings of the National Academy of Sciences of the United States of America* 91, 5720–5724.
- Squire, L.R., 1992. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychology Review* 99, 195–231.
- Stark, C.E., 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.
- Viskontas, I.V., Knowlton, B.J., Steinmetz, P.N., Fried, I., 2006. Differences in mnemonic processing by neurons in the human hippocampus and parahippocampal regions. *Journal of Cognitive Neuroscience* 18, 1654–1662.
- Waldron, E.M., Ashby, F.G., 2001. The effects of concurrent task interference on category learning: evidence for multiple category learning systems. *Psychonomic Bulletin & Review* 8, 168–176.
- Wickens, J., 1990. Striatal dopamine in motor activation and reward-mediated learning: steps towards a unifying model. *Journal of Neural Transmission-General Section* 80, 9–31.
- Willingham, D.B., 1998. A neuropsychological theory of motor skill learning. *Psychology Review* 105, 558–584.
- Willingham, D.B., Wells, L.A., Farrell, J.M., Stemwedel, M.E., 2000. Implicit motor sequence learning is represented in response locations. *Memory and Cognition* 28, 366–375.
- Wilson, C., 1995. *The Contribution of Cortical Neurons to the Firing Pattern of Striatal Spiny Neurons*. Bradford, Cambridge, MA.
- Yeterian, E.H., Pandya, D.N., 1995. Corticostriatal connections of extrastriate visual areas in rhesus monkeys. *Journal of Comparative Neurology* 352, 436–457.
- Zeithamova, D., Maddox, W.T., 2006. Dual-task interference in perceptual category learning. *Memory and Cognition* 34, 387–398.