in: L. Jiménez (Ed.), *Attention and implicit learning* (pp. 109-141). Amsterdam: John Benjamins Publishing Company.

The cognitive neuroscience of implicit category learning

F. Gregory Ashby & Michael B. Casale*
University of California, Santa Barbara

There is much recent interest in the question of whether people have available a single category learning system or a number of qualitatively different systems. Most proponents of multiple systems have hypothesized an explicit, rule-based system and some type of implicit system. Although there has been general agreement about the nature of the explicit system, there has been disagreement about the exact nature of the implicit system. This chapter explores the question of whether there is implicit category learning, and if there is, what form it might take. First, we examine what the word "implicit" means in the categorization literature. Next, we review some of the evidence that supports the notion that people have available one or more implicit categorization systems. Finally, we consider the nature of implicit categorization by focusing on three alternatives: an exemplar memory-based system, a procedural memory system, and an implicit system that uses the perceptual representation memory system.

1. The cognitive neuroscience of implicit category learning

Categorization is the act of responding differently to objects and events in the environment that belong to separate classes or categories. It is a critical process that every organism must perform in at least a rudimentary form because it allows them to respond differently, for example, to nutrients and poisons, and to predators and prey.

Much of the recent categorization literature has focused on the question of whether people have available a single category learning system or a number of qualitatively different systems. For example, although the early literature

was dominated by theories postulating a single system, a number of recent theories have proposed multiple category learning systems (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Brooks, 1978; Erickson & Kruschke, 1998; Pickering, 1997). Interestingly, many of these papers have hypothesized at least two similar systems: 1) an explicit, rule-based system that is tied to language function and conscious awareness, and 2) an implicit system that may not have access to conscious awareness. For example, Ashby et al. (1998) proposed a formal neuropsychological theory of multiple category learning systems called COVIS (COmpetition between Verbal and Implicit Systems), which assumes separate explicit (rule-based) and implicit (procedural learning-based) systems.

There is still much disagreement however. First, the proposal that there are multiple category learning systems is disputed. In particular, Nosofsky and his colleagues have argued that single system models can account for many of the phenomena that have been used to support the notion of multiple systems (Nosofsky & Johansen, 2000; Nosofsky & Kruschke, 2002; Nosofsky & Zaki, 1998). Second, even among those researchers postulating separate explicit and implicit systems, there is disagreement about the nature of the implicit system. As mentioned above, Ashby et al. (1998) proposed a procedural-memory based implicit system (see also Ashby & Waldron, 1999; Ashby, Waldron, Lee, & Berkman, 2001). In contrast, several researchers have proposed that the implicit system is exemplar-memory based (Erickson & Kruschke, 1998; Pickering, 1997), and there have also been proposals that the perceptual representation memory system participates in implicit category learning (Ashby & Ell, 2001; Knowlton, Squire et al., 1996; Reber, Stark, & Squire, 1998).

This chapter explores the question of whether there is implicit category learning. First, we examine what is meant by explicit and implicit categorization. These are important questions because both terms are used somewhat differently in the categorization literature than in the memory literature. Next, we briefly review evidence supporting the notion that people have available one or more implicit categorization systems. Finally, we focus on two putative implicit category learning systems, one that uses procedural memory and one that uses the perceptual representation memory system.

2. What are explicit and implicit categorization?

2.1 Explicit categorization

Categorization processes are said to be explicit if they are accessible to conscious awareness. This would include traditional declarative memory processes that might be invoked when participants try to memorize responses associated with the various stimuli. However, it could also include simple rule-based strategies such as, "the stimulus belongs to category A if it is red, and it belongs to category B if it is blue."

One danger with equating explicit processing with conscious awareness is that this shifts the debate from how to define 'explicit' to how to define 'conscious awareness'. Ashby et al. (1998) suggested that one pragmatic solution to this problem is to adopt the criterion that category learning is explicit if the subject can verbally describe the categorization rule that he or she used. This definition works well in most cases, but it seems unlikely that verbalizability should be a requirement for explicit reasoning. For example, the insight displayed by Köhler's (1925) famous apes seems an obvious example of explicit reasoning in the absence of language. For now we will use the criterion of verbalizability for explicit category learning but ultimately, a theoretically motivated criterion for conscious awareness is needed.

One way to develop a theory of conscious awareness is by exploiting the relationship between awareness and working memory. For example, the contents of working memory are clearly accessible to conscious awareness. In fact, because of its close association to executive attention, a strong argument can be made that the contents of working memory *define* our conscious awareness. When we say that we are consciously aware of some object or event, we mean that our executive attention has been directed to that stimulus. Its representation in our working memory gives it a moment-to-moment permanence. Working memory makes it possible to link events in the immediate past with those in the present, and it allows us to anticipate events in the near future. All of these are defining properties of conscious awareness.

The association between working memory and the prefrontal cortex makes it possible to formulate cognitive neuroscience models of conscious awareness. The most influential such model was developed by Francis Crick and Christof Koch (Crick & Koch, 1990, 1995, 1998). The Crick-Koch hypothesis states that one can have conscious awareness only of activity in brain areas that project directly to the prefrontal cortex[1]. For example, consider two brain areas X and Y. Suppose area X projects directly to the prefrontal cortex, but area Y

projects only to area X (i.e., and not directly to the prefrontal cortex). If working memory and conscious awareness reside in prefrontal cortex, then we can be consciously aware of activity in area X because it can be loaded directly into working memory. On the other hand, if activity in area Y is transformed by area X before reaching prefrontal cortex and conscious awareness, then there is no way to be aware of activity in area Y – only of the transformed activity that leaves area X.

Primary visual cortex (Area V1) does not project directly to the prefrontal cortex, so the Crick-Koch hypothesis asserts that we cannot be consciously aware of activity in V1. Crick and Koch (1995, 1998) described evidence in support of this prediction. Of course, many other brain regions also do not project directly to the prefrontal cortex. For example, the basal ganglia do not project directly to the prefrontal cortex (i.e., they first project through the thalamus), so the Crick-Koch hypothesis predicts that we are not aware of activity within the basal ganglia. Memory theorists believe that the basal ganglia mediate procedural memories (Jahanshahi, Brown, & Marsden, 1992; Mishkin, Malamut, & Bachevalier, 1984; Saint-Cyr, Taylor, & Lang, 1988; Willingham, Nissen, & Bullemer, 1989), so the Crick-Koch hypothesis provides an explanation of why we don't seem to be aware of procedural (e.g., motor) learning.

In summary, although the Crick-Koch hypothesis offers a promising start, a complete theory of conscious awareness does not yet exist. Therefore, in this chapter we will adopt the operational definition that a categorization process is explicit if it can be described verbally.

2.2 Implicit categorization

During the past 10 years, about 120 articles have appeared in the psychological literature that discuss implicit category learning or implicit categorization, whereas the decade of the 1980s saw only about 20 such articles. This recent interest in implicit category learning has profoundly affected the categorization literature, and has formed bridges to the memory literature, where of course, the study of implicit processes have a long and rich history. Even so, a memory researcher interested in implicit categorization may be confused by how the term "implicit" is used in the categorization literature.

Many memory theorists adopt the strong criterion that a memory is implicit only if there is no conscious awareness of its details *and* there is no knowledge that a memory has even been stored (e.g., Schacter, 1987). In a typical categorization task, these criteria are impossible to meet because trial-by-trial feedback is routinely provided. When an observer receives feedback that a

response is correct, then this alone makes it obvious that learning has occurred, even if there is no internal access to the system that is mediating this learning. Thus, in category learning, a weaker criterion for implicit learning is typically used in which the observer is required only to have no conscious access to the nature of the learning, even though he or she would be expected to know that some learning has occurred.

The stronger criterion for implicit processing that has been adopted in much of the memory literature could be applied in unsupervised category learning tasks, in which no trial-by-trial feedback of any kind is provided. In the typical unsupervised task, observers are told the number of contrasting categories and are asked to assign stimuli to these categories, but are never told whether a particular response is correct or incorrect. Free sorting is a similar, but more unstructured task in which participants are not given feedback about the accuracy of their responses, nor are they even told the number of contrasting categories (e.g., Ashby & Maddox, 1998). Thus, in both unsupervised and free sorting tasks there is no feedback that observers can use to infer that learning has occurred. As a result, these tasks are ideal for using the stricter criterion to test for implicit learning. Even so, to date the only learning that has been demonstrated in such tasks is explicit (Ashby, Queller, & Berretty, 1999; Medin, Wattenmaker, & Hampson, 1997).

3. Evidence for separate explicit and implicit category learning systems

3.1 Three different category learning tasks

Much of the data that has been used to argue for multiple category learning systems came from the observation that changing the nature of the contrasting categories that subjects were asked to learn sometimes qualitatively changed learning behavior. Ashby and Ell (2001) identified three different types of category structures that are often associated with such qualitative differences in performance. To anticipate our later discussion, in the next section we will argue that these three tasks load primarily on three different memory systems.

Rule-based tasks are those in which subjects can learn the category structures via some explicit reasoning process. In the most common applications, only one stimulus dimension is relevant, and the subject's task is to discover this relevant dimension and then to map the different dimensional values to the relevant categories. Figure 1 shows the stimuli and category structure of a recent rule-based task that used 8 exemplars per category

(Waldron & Ashby, 2001). The categorization stimuli were colored geometric figures presented on a colored background. The stimuli varied on four binary-valued dimensions: background color (blue or yellow; here denoted as light and dark gray, respectively), embedded symbol color (green or red; here denoted as black and white, respectively), symbol number (1 or 2), and symbol shape (square or circle). This yields a total of 16 possible stimuli. To create rule-based category structures, one dimension is selected arbitrarily to be relevant. The two values on that dimension are then assigned to the two contrasting categories.

An important property of rule-based category learning tasks is that the optimal rule is often easy to describe verbally (Ashby et al., 1998). As a result, subjects can learn the category structures via an explicit process of hypothesis testing (Bruner, Goodnow, & Austin, 1956) or theory construction and testing (Murphy & Medin, 1985). Virtually all standard neuropsychological categorization tasks are of this type – including the well known Wisconsin Card Sorting Test (Heaton, 1981). Rule-based tasks, which have a long history in cognitive psychology, have been favored by proponents of the so-called classical theory of categorization, which assumes that category learning is the process of discovering the set of necessary and sufficient conditions that determine category membership (Smith & Medin, 1981).

In the Figure 1 example, the explicit rule that perfectly separates the stimuli into the two categories is unidimensional. Although the optimal rule in rule-based tasks is often unidimensional, this is not a requirement. For example, a task is rule-based if the optimal rule is a conjunction of the form:

Respond A if the background is blue and the embedded symbol is round;
otherwise respond B.

The critical criterion is that this rule is easy to describe verbally, and to learn through an explicit reasoning process. Note that according to this criterion, there is no limit on the complexity of the optimal rule in rule-based tasks. However, as the complexity of the optimal rule increases, its salience decreases and it becomes less likely that observers will learn the associated categories through an explicit reasoning process. In fact, Alfonso-Reese (1997) found that even simple conjunction rules have far lower salience than unidimensional rules. This does not mean that people can not learn conjunction rules. Only that they are unlikely to experiment with such rules unless feedback compels them in this direction. This discussion should make it clear that the boundary on what constitutes a rule-based task is fuzzy. Tasks in which the optimal rule is unidimensional are unambiguously rule-based (at least with separable stimulus dimensions), and tasks in which the optimal rule is significantly more complex than a conjunction rule are almost never rule-based. In between, the

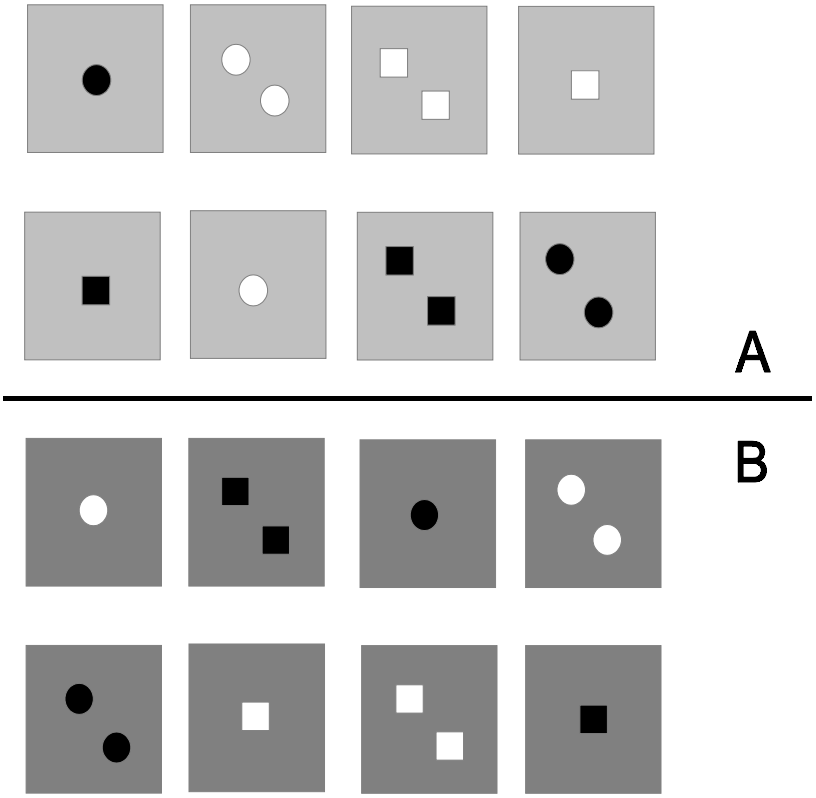

Figure 1. Category structure of a rule-based category learning task. The optimal explicit rule is: Respond A if the background color is blue (depicted as light gray), and respond B if the background color is yellow (depicted as dark gray).

classification is not so clear-cut. For this reason, the rule-based tasks we discuss in this chapter will all have a unidimensional optimal rule.

Information-integration tasks are those in which accuracy is maximized only if information from two or more stimulus components (or dimensions) is integrated at some pre-decisional stage (Ashby & Gott, 1988). Perceptual integration could take many forms – from treating the stimulus as a Gestalt to computing a weighted linear combination of the dimensional values. However, a conjunction rule is a rule-based task rather than an information-integration task because separate decisions are first made about each dimension (e.g., small or large) and then the outcome of these decisions is combined (integration is not pre-decisional). In many cases, the optimal rule in information-integration tasks is difficult or impossible to describe verbally (Ashby et al., 1998).

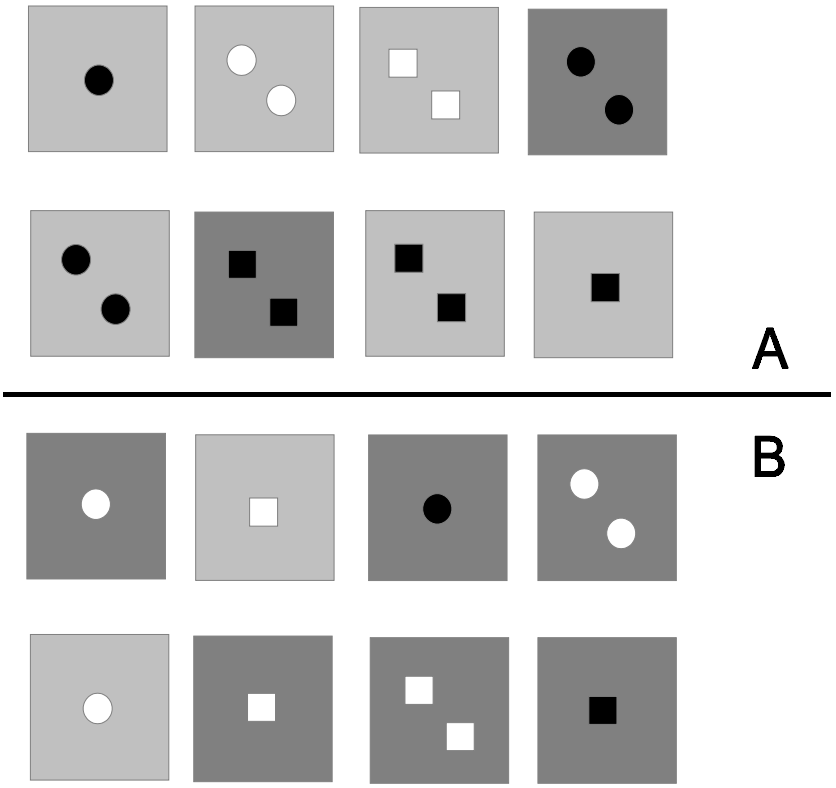

Figure 2. Category structure of an information integration category learning task with only a few exemplars in each category.

The neuropsychological data reviewed below suggests that performance in such tasks is qualitatively different depending on the size of the categories – in particular, when a category contains only a few highly distinct exemplars, memorization is feasible. However, when the relevant categories contain many exemplars (e.g., hundreds), memorization is less efficient.

Figure 2 shows the stimuli and category structure of a recent information-integration task that used only 8 exemplars per category (Waldron & Ashby, 2001). The categorization stimuli are the same as in Figure 1. To create information-integration category structures, one dimension is arbitrarily selected to be irrelevant. For example, in Figure 2, the irrelevant dimension is symbol shape. Next, one level on each relevant dimension is arbitrarily assigned a value of +1 and the other level is assigned a value of 0. In Figure 2, a background color of blue (denoted as light gray), a symbol color of green

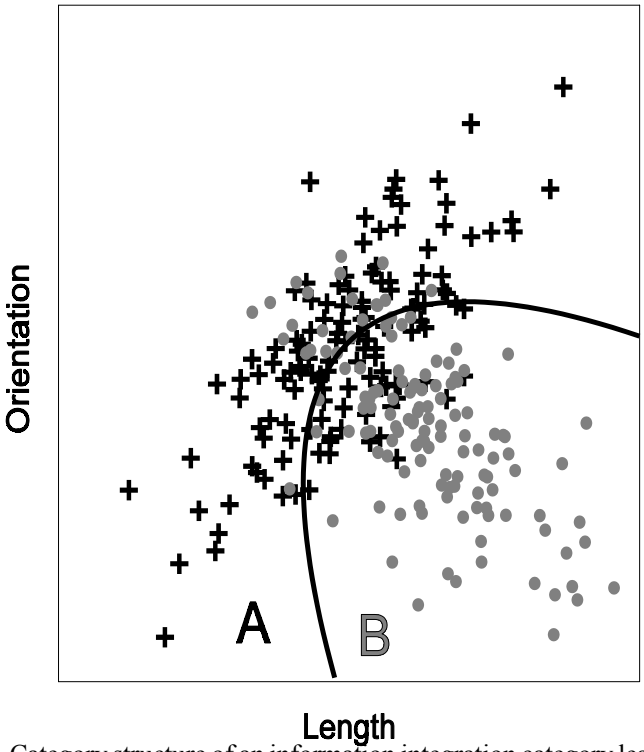

Figure 3. Category structure of an information integration category learning task with many exemplars per category. Each stimulus is a line that varies across trials in length and orientation. Every black plus depicts the length and orientation of a line in Category A and every gray dot depicts the length and orientation of a line in Category B. The quadratic curve is the boundary that maximizes accuracy.

(denoted as black), and a symbol number of 2 are all assigned a value of +1. Finally, the category assignments are determined by the following rule:

The stimulus belongs to category A if the sum of values on the relevant dimensions > 1.5; Otherwise it belongs to category B.

This rule is readily learned by healthy young adults, but even after achieving perfect performance, they can virtually never accurately describe the rule they used[2].

Figure 3 is an abstract representation of the category structure of an information-integration task in which there are hundreds of exemplars in each category (developed by Ashby & Gott, 1988). In this experiment, each stimulus is a line that varies across trials in length and orientation. Each cross in Figure 3 denotes the length and orientation of an exemplar in Category A and each dot

denotes the length and orientation of an exemplar in Category B**.** The categories overlap, so perfect accuracy is impossible in this example. Even so, the quadratic curve is the boundary that maximizes response accuracy – that is, accuracy is maximized if subjects respond B to any stimulus falling inside the quadratic region (in the lower right quadrant), and A to any stimulus falling outside of this region. Note that such a rule is impossible to describe verbally. Many experiments have shown that, given enough practice, the performance of subjects in this task is well described by a quadratic decision boundary (e.g., Ashby & Maddox, 1992; Maddox & Ashby, 1993).

Information-integration tasks with few exemplars per category have been the favorites of exemplar theorists, who argue that categorization requires accessing the memory representations of every previously seen exemplar from each relevant category (e.g., Estes, 1986; 1994; Medin & Schaffer, 1978; Nosofsky, 1986; Smith & Minda, 2000). In contrast, decision bound theorists, who argue that category learning is a process of associating category labels with regions of perceptual space, have traditionally used information-integration tasks with many exemplars per category (e.g., Ashby & Maddox, 1992; Maddox & Ashby, 1993).

Prototype distortion tasks are a third type of category learning task in which each category is created by first defining a category prototype and then creating the category members by randomly distorting these prototypes. In the most popular version of the prototype distortion task, the category exemplars are random dot patterns (Posner & Keele, 1968). An example is shown in Figure 4. In a typical application, many stimuli are created by randomly placing a number of dots on the display. One of these dot patterns is then chosen as the prototype for category A. The others become stimuli not belonging to category A. The other exemplars in category A are then created by randomly perturbing the position of each dot in the category A prototype. A consequence of this process that will prove important in our later discussions is that the stimuli that are not in category A have no coherent structure. For this reason, participants are often instructed to respond "yes" or "no" depending on whether the presented stimulus is a member of category A, rather than "A" or "B" as in the tasks illustrated in Figures 1 - 3. As the name suggests, prototype distortion tasks have been commonly used by prototype theorists, who argue that categorization is the act of comparing the presented stimulus to the prototype of each contrasting category (Homa, Sterling, & Trepel, 1981; Posner & Keele, 1968; Minda & Smith, 2001).

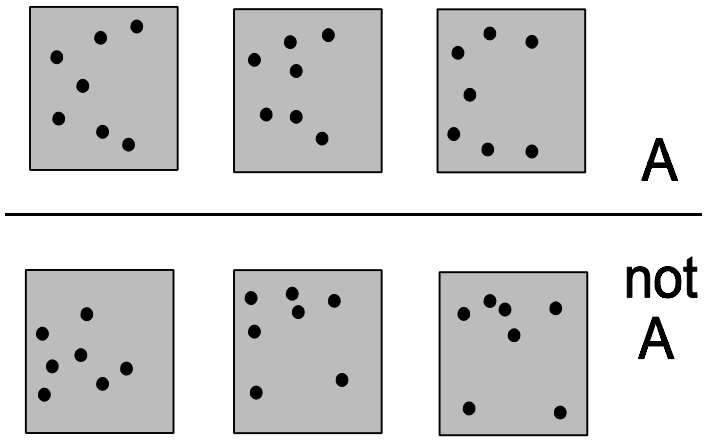

Figure 4. Some exemplars from a prototype distortion category learning task with random dot patterns.

3.2 Category learning dissociations

We have now observed a number of different dissociations between performance in rule-based and information-integration category learning tasks. Collectively, these provide strong evidence that learning in these two types of tasks is mediated by separate systems. A number of these results show that the nature and timing of trial-by-trial feedback about response accuracy is critical with information-integration categories but not with rule-based categories. First, in the absence of any trial-by-trial feedback about response accuracy, people can learn some rule-based categories, but there is no evidence that they can learn information-integration categories (Ashby, Queller, & Berretty, 1999). Second, even when feedback is provided on every trial, information-integration category learning is impaired if the feedback signal is delayed by as little as five seconds after the response. In contrast, such delays have no effect on rule-based category learning (Maddox, Ashby, & Bohil, 2002). Third, similar results are obtained when observational learning is compared to traditional feedback learning. Ashby, Maddox, and Bohil (2002) trained subjects on rule-based and information-integration categories using an observational training paradigm in which subjects are informed before stimulus presentation of what category the ensuing stimulus is from. Following stimulus presentation, subjects then pressed the appropriate response key. Traditional feedback training was as effective as observational training with rule-based

categories, but with information-integration categories, feedback training was significantly more effective than observational training.

Another qualitative difference between these two tasks is that information-integration category learning is more closely tied to motor outputs than rule-based category learning. Ashby, Ell, and Waldron (2002) had subjects learn either rule-based or information integration categories using traditional feedback training. Next, some subjects continued as before, some switched their hands on the response keys, and for some the location of the response keys was switched (so the Category A key was assigned to Category B and vice versa). For those subjects learning rule-based categories, there was no difference among any of these transfer instructions, thereby suggesting that abstract category labels are learned in rule-based categorization. In contrast, for those subjects learning information-integration categories, switching hands on the response keys caused no interference, but switching the locations of the response keys caused a significant decrease in accuracy. Thus, it appears that response locations are learned in information-integration categorization, but not specific motor programs.

One criticism of all these results is that information-integration tasks are usually more difficult than rule-based tasks, in the sense that information integration tasks usually require more training to reach the same level of expertise. Because of this difficulty difference, one concern is that, collectively, these studies might show only that there are many ways to disrupt learning of difficult tasks compared to simpler tasks. However, several results argue strongly against this hypothesis. First, Waldron and Ashby (2001) had subjects learn rule-based and information-integration categories (shown in Figures 1 and 2, respectively) under typical single-task conditions and when simultaneously performing a secondary task known to activate frontal cortical structures (i.e., a numerical Stroop task). If task difficulty was the relevant variable, then the dual task should interfere more strongly with the difficult information-integration task than with the simpler rule-based task (since it is harder to do two difficult things at once than two simple things). However, in contrast to this prediction, the dual task interfered much more strongly with the ability of subjects to learn the rule-based task than the information-integration task.

Second, Ashby, Noble et al. (2002) found that the same group of Parkinson's disease patients were much more impaired at rule-based category learning (the Figure 1 task) than at information integration category learning (the Figure 2 task). If a single system mediates learning in these two types of categorization tasks, and if Parkinson's disease damages this system, then we

would expect the more serious deficits to occur in the more difficult information integration tasks.

These dissociations strongly argue that people learn rule-based and information-integration categories using separate systems. For example, consider just the single Waldron and Ashby (2001) dual-task experiment. Arguably the most successful existing single-process model of category learning is Kruschke's (1992) ALCOVE model. Ashby and Ell (2002a) showed that the only versions of ALCOVE that can fit the Waldron and Ashby data make the strong prediction that after reaching criterion accuracy on the simple (unidimensional) rule-based structures, participants would have no idea that only one dimension was relevant in the dual-task conditions. Ashby and Ell reported empirical evidence that strongly disconfirmed this prediction of ALCOVE. Thus, the best available single-system model fails to account even for the one dissociation reported by Waldron and Ashby (2001).

In addition to dissociations in experiments with healthy young adults, a number of related dissociations have been reported with neuropsychological patient groups. In particular, Ashby and Ell (2001) reviewed the current neuropsychological category learning data and found evidence of a different set of dissociations across these three categorization tasks. Presently, there is extensive category learning data on only a few neuropsychological populations. The best data come from four different groups: 1) patients with frontal lobe lesions, 2) patients with medial temporal lobe amnesia, and two types of patients suffering from a disease of the basal ganglia – either 3) Parkinson's or 4) Huntington's disease. Table 1 summarizes the performance of these groups on the three different types of category learning tasks.

Note first that Table 1 indicates a double dissociation between frontal lobe patients and medial temporal lobe amnesiacs on rule-based tasks and information-integration tasks with few exemplars per category. Specifically, frontal patients are impaired on rule-based tasks (e.g., the Wisconsin Card Sorting Test; Kolb & Whishaw, 1990) but medial temporal lobe amnesiacs are normal (e.g., Janowsky, Kritchevsky, & Squire, 1989; Leng & Parkin, 1988). At the same time, the available data on information-integration tasks with few exemplars per category indicates that frontal patients are normal (Knowlton, Mangels, & Squire, 1996), but medial temporal lobe amnesiacs are impaired (i.e., they show a late-training deficit -- that is, they learn normally during the first 50 trials or so, but thereafter show impaired learning relative to age-matched controls; Knowlton, Squire, & Gluck, 1994). Therefore, the neuropsychological data also support the hypothesis that at least two systems

Table 1. Performance of Various Neuropsychological Populations on Three Types of Category Learning Tasks.

Neuropsychological Group		Task			
		Rule-Based	Information-Integration		Prototype Distortion
			Many Exemplars	Few Exemplars	
Frontal Lobe Lesions		Impaired	?	Normal	?
Basal Ganglia Disease	Parkinson's Disease	Impaired	Impaired	Impaired	?
	Huntington's Disease	Impaired	Impaired	Impaired	?
Medial Temporal Lobe Amnesia		Normal	Normal	Late Training Deficit	Normal

participate in category learning. Of course, until more data are collected on the information-integration tasks, this conclusion must be considered tentative.

Table 1 can also be used to construct first hypotheses about which neural structures mediate learning in the various category learning tasks. For example, patients with frontal or basal ganglia dysfunction are impaired in rule-based tasks (e.g., Brown & Marsden, 1988; Cools et al.,1984; Kolb & Whishaw, 1990; Robinson, Heaton, Lehman, & Stilson, 1980), but patients with medial temporal lobe damage are normal in this type of category learning task (e.g., Janowsky et al., 1989; Leng & Parkin, 1988). Thus, an obvious first hypothesis is that the prefrontal cortex and the basal ganglia participate in this type of learning, but the medial temporal lobes do not. Converging evidence for the hypothesis that these are important structures in rule-based category learning comes from several sources. First, an fMRI study of a rule-based task similar to the Wisconsin Card Sorting Test showed activation in the right dorsal-lateral prefrontal cortex, the anterior cingulate, and the head of the right caudate nucleus (among other regions) (Rao et al., 1997). Similar results were recently obtained in an fMRI study of the Wisconsin Card Sorting Test (Monchi et al., 2001). Second, many studies have implicated these structures as key components of executive attention (Posner & Petersen, 1990) and working memory (e.g., Fuster, 1989; Goldman-Rakic, 1987, 1995), both of which are likely to be critically important to the explicit processes of rule formation and testing that are assumed to mediate rule-based category learning. Third, a recent neuroimaging study identified the (dorsal) anterior cingulate as the site of hypothesis generation in a rule-based category-learning task (Elliott &

Dolan, 1998). Fourth, lesion studies in rats implicate the dorsal caudate nucleus in rule switching (Winocur & Eskes, 1998).

Next, note that in information integration tasks with large categories, only patients with basal ganglia dysfunction are known to be impaired (Filoteo, Maddox, & Davis, 2001a; Maddox & Filoteo, 2001). In particular, medial temporal lobe patients are normal (Filoteo, Maddox, & Davis, 2001b). So a first hypothesis should be that the basal ganglia are critical in this task, but the medial temporal lobes are not. If the number of exemplars per category is reduced in this task to a small number (e.g., 4 to 8), then medial temporal lobe amnesiacs show late training deficits – that is, they learn normally during the first 50 trials or so, but thereafter show impaired learning relative to age-matched controls (Knowlton, Squire, & Gluck, 1994). An obvious possibility in this case, is that normal observers begin memorizing responses to at least a few of the more distinctive stimuli – a strategy that is not available to the medial temporal lobe amnesiacs, and which is either not helpful or impossible when the categories contain many exemplars. Since patients with basal ganglia dysfunction are also impaired with small categories requiring information-integration (Knowlton, Mangels et al., 1996; Knowlton, Squire et al., 1996), a first hypothesis should be that learning in such tasks depends on the basal ganglia and on medial temporal lobe structures.

Finally, to our knowledge, of the patient groups identified in Table 1, only amnesiacs have been run in prototype distortion tasks. Several studies have reported that this patient group shows normal learning in prototype distortion tasks, which suggests that learning in this task does not depend on an intact medial temporal lobe (Knowlton & Squire, 1993; Kolodny, 1994). Ashby and Ell (2001) suggested that under certain conditions, learning in prototype distortion tasks might depend, in part, on the perceptual representation memory system – through a perceptual learning process. In the random dot pattern experiments, this seems plausible because all category A exemplars are created by randomly perturbing the positions of the dots that form the category A prototype (see Figure 4). Thus, if there are cells in visual cortex that respond strongly to the category A prototype, they are also likely to respond to the other category A exemplars, and perceptual learning will increase their response. If this occurs, the observer could perform well in this task by responding "yes" to any stimulus that elicits a strong feeling of visual familiarity. Recent fMRI studies of subjects in prototype distortion tasks show learning related changes in visual cortex (Reber et al., 1998), and are thus consistent with this hypothesis. Before drawing any strong conclusions however, it is vital to obtain

category learning data on prototype distortion tasks from patients with basal ganglia disease or frontal lobe lesions.

In artificial grammar learning, subjects must decide whether or not a letter string has a familiar (artificial) grammatical structure (e.g., Reber, 1989). Although seemingly very different from prototype distortion, it has also been proposed that artificial grammar learning depends on the perceptual representation memory system (Knowlton, Squire et al., 1992). Indirect support for this hypothesis comes from a number of studies showing that amnesiacs and basal ganglia disease patients exhibit normal artificial grammar learning (Knowlton, Squire et al., 1996; Knowlton, Ramus, & Squire, 1992; Meulemans, Peigneux, & Van der Linden, 1998). Future research should explore the possible connections between prototype distortion category learning and artificial grammar learning.

4. Are there multiple implicit category learning systems?

The results reviewed above suggest that there may be multiple qualitatively different implicit category learning systems. Two obvious possibilities are a procedural-learning based system that is mediated, in part, by the basal ganglia, and a perceptual representation system that relies on perceptual learning in visual cortex. The next two sections consider these possibilities in some detail. A third possibility that should also be considered, however, is whether there is an exemplar memory-based system.

In cognitive psychology, one of the most popular and influential theories of category learning is exemplar theory (Brooks, 1978; Estes, 1986; Medin & Schaffer, 1978; Nosofsky, 1986), which assumes that categorization decisions are made by accessing memory representations of all previously seen exemplars. Exemplar theorists are careful not to assume that this process of accessing memory representations is explicit, but most exemplar theorists have not taken a strong stand about the neural basis by which these memory representations are encoded. A natural candidate is the hippocampus and other medial temporal lobe structures (e.g., Pickering, 1997). However, this is problematic because these brain areas are thought to mediate (the consolidation of) episodic memory, which is considered to be explicit (Fuster, 1989; Knowlton & Squire, 1993; Reber & Squire, 1994). Certainly people are not consciously aware of recalling all previously seen exemplars when making categorization decisions.

There are situations in which episodic memory may contribute to category learning. In particular, with categories that contain a few highly distinct

exemplars, people may memorize responses to at least some category members. Then, when a particularly distinct exemplar is presented, subjects may use episodic memory to recall the correct response. As mentioned previously, this might be the cause of the late-training deficit that has been reported when medial temporal lobe amnesiacs learn information-integration categories. Note, however, that the possible use of episodic memory to recall the response associated with the single current stimulus is very different from the processes hypothesized by exemplar theory. According to exemplar theory, the memory representations of all previously seen exemplars are accessed on every trial. Although they both seem to involve a similar type of memory trace, psychologically these two possibilities are very different. Recalling the response to a distinct stimulus is an explicit process, whereas accessing all previously seen exemplars almost necessarily must be implicit (since subjects report no awareness of such massive activation). On the other hand, there is evidence that at least some of the success of exemplar theory is due to the ability of exemplar models to mimic this explicit recall process. For example, Smith and Minda (2000) found that the best fits of a powerful exemplar model to category learning data collected using a popular information integration category structure (with a few highly distinct stimuli in each category) occurred when the response probabilities were determined almost completely by the presented stimulus. The representations of other category members were also activated, but the model parameters were such that these were so dissimilar to the presented stimulus that they had virtually no effect on the predictions of the model.

In summary, there is some evidence that an explicit, episodic memory-based process may contribute to category learning in some situations (e.g., when categories contain a few highly distinct exemplars). There is also theoretical reason to expect that an implicit exemplar memory-based system may contribute to category learning. However, the only attempts that have been made to describe the neurobiological basis of such a system have focused on the hippocampus and related structures that are thought to mediate explicit, episodic memories (Gluck, Oliver, & Myers, 1996; Pickering, 1997). Thus, currently, an unresolved, but extremely important question is whether there exists some implicit, exemplar-memory based categorization system.

5. A procedural learning-based categorization system

Figure 5 shows the circuit of a putative procedural memory-based category learning system (proposed by Ashby et al.,1998; Ashby & Waldron, 1999). The

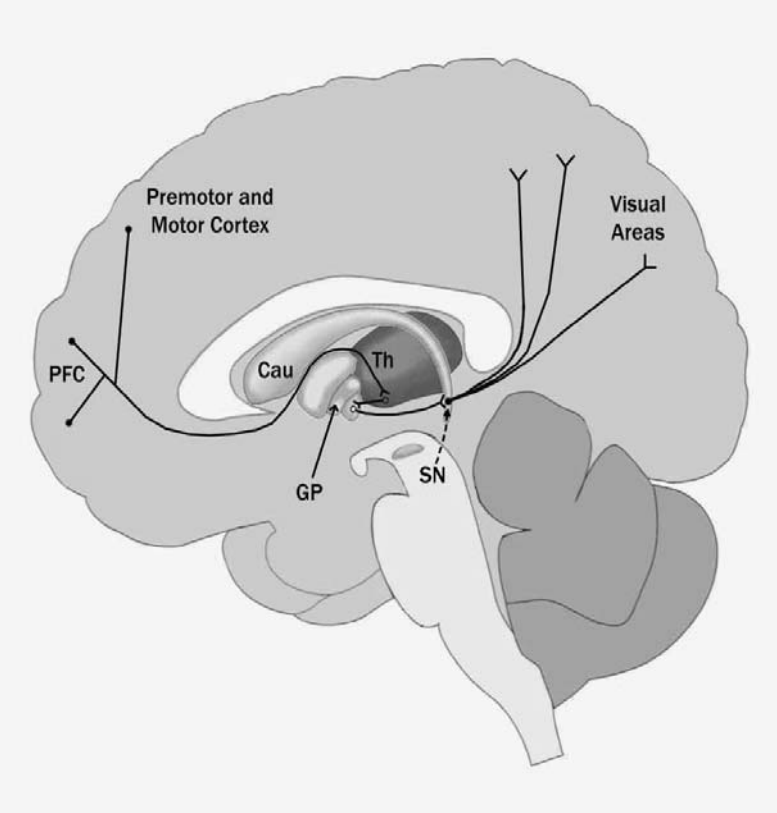

Figure 5. A procedural-memory-based category learning system. Excitatory projections end in solid circles, inhibitory projections end in open circles, and dopaminergic projections are dashed. PFC = prefrontal cortex, Cau = caudate nucleus, GP = globus pallidus, and Th = thalamus.

key structure in this model is the caudate nucleus, a major input structure within the basal ganglia. In primates, all of extrastriate visual cortex projects directly to the tail of the caudate nucleus, with about 10,000 visual cortical cells converging on each caudate cell (Wilson, 1995). Cells in the tail of the caudate (i.e., medium spiny cells) then project to prefrontal and premotor cortex (via the globus pallidus and thalamus; e.g., Alexander, DeLong, & Strick, 1986). The model assumes that, through a procedural learning process, each caudate unit learns to associate a category label, or perhaps an abstract motor program, with a large group of visual cortical cells (i.e., all that project to it).

Perhaps the best evidence for a basal ganglia contribution to category learning comes from a long series of lesion studies in rats and monkeys that

show that the tail of the caudate nucleus is both necessary *and* sufficient for visual discrimination learning. Many studies have shown that lesions of the tail of the caudate nucleus impair the ability of animals to learn visual discriminations that require one response to one stimulus and a different response to some other stimulus (e.g., McDonald & White, 1993, 1994; Packard, Hirsch, & White, 1989; Packard & McGaugh, 1992). For example, in one study, rats with lesions in the tail of the caudate could not learn to discriminate between safe and unsafe platforms in the Morris water maze when the safe platform was marked with horizontal lines and the unsafe platform was marked with vertical lines (Packard & McGaugh, 1992). The same animals learned normally, however, when the cues signaling which platform was safe were spatial. Since the visual cortex is intact in these animals, it is unlikely that their difficulty is in perceiving the stimuli. Rather, it appears that their difficulty is in learning to associate an appropriate response with each stimulus alternative, and in fact, many researchers have hypothesized that this is the primary role of the neostriatum (e.g., Rolls, 1994; Wickens, 1993). Technically, such studies are categorization tasks with one exemplar per category. It is difficult to imagine how adding more exemplars to each category could alleviate the deficits caused by caudate lesions, and it is for this reason that the caudate lesion studies support the hypothesis that the caudate contributes to normal category learning.

The sufficiency of the caudate nucleus for visual discrimination learning was shown in a series of studies by Gaffan and colleagues that lesioned all pathways out of visual cortex except into the tail of the caudate (e.g., projections into prefrontal cortex were lesioned by Eacott & Gaffan, 1991, and Gaffan & Eacott, 1995; projections to the hippocampus and amygdala were lesioned by Gaffan & Harrison, 1987). None of these lesions affected visual discrimination learning.

The procedural learning that has been hypothesized to occur in the caudate nucleus is thought to be facilitated by a dopamine mediated reward signal from the substantia nigra (pars compacta) (e.g., Wickens, 1993). There is a large literature linking dopamine and reward, and many researchers have argued that a primary function of dopamine is to serve as the reward signal in reward-mediated learning (e.g., Beninger, 1983; Miller, Sanghera, & German, 1981; Montague, Dayan, & Sejnowski, 1996; White, 1989; Wickens, 1993). For example, it has been shown that rewards, and events that signal reward, elicit release of dopamine from several brainstem sites (for reviews, see, e.g, Bozarth, 1994; Pfaus & Phillips, 1991; Philips, Blaha, Pfaus, & Blackburn, 1992), and it is well known that dopamine antagonists (i.e., neuroleptics) disrupt the reward signal and render reinforcement ineffective (e.g., Ataly & Wise, 1983).

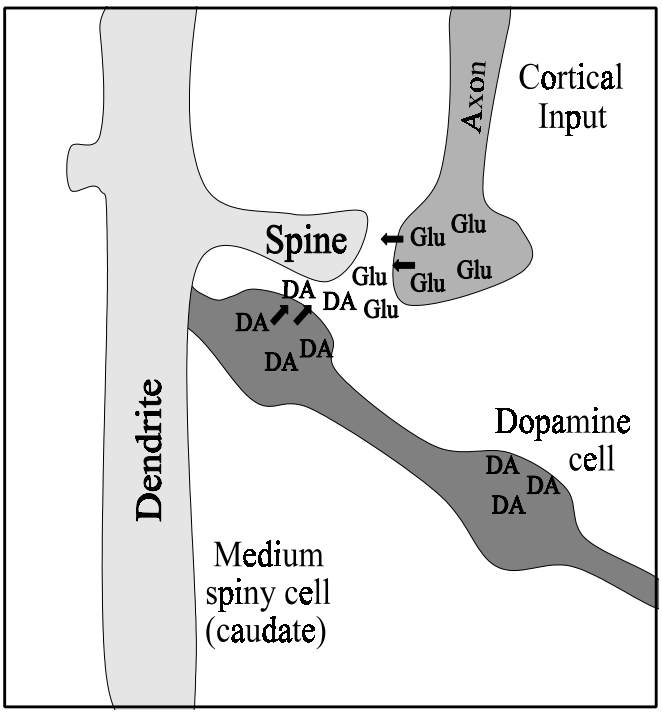

Figure 6. A closer view of a cortical-striatal synapse. Here, a cortical cell terminal releases glutamate (Glu) onto the dendritic spine of a medium spiny cell of the caudate nucleus. Dopamine cells of the substantia nigra also project onto medium spiny cells and upon presentation of reward, release dopamine (DA) into the same synapse.

Fairly specific neurobiological models of this learning process have been developed (e.g., Wickens, 1993). Figure 6 shows a close-up view of a synapse between the axon of a pyramidal cell originating in visual cortex and the dendrite of a medium spiny cell in the caudate nucleus. Note that glutamate projections from visual cortex and dopamine projections from the substantia nigra both synapse on the dendritic spines of caudate medium spiny cells (DiFiglia, Pasik, & Pasik, 1978; Freund, Powell, & Smith, 1984; Smiley et al., 1994). A cortical signal causes an influx of free Ca^{2+} into the spines (through NMDA receptors). Because of its strong positive charge, free Ca^{2+} is buffered very quickly within the intracellular medium. The main effect of Ca^{2+} entering the cell is to activate Ca-dependent protein kinases, which then perform a number of cellular functions, including strengthening (long term potentiation -- LTP) and weakening (long term depression -- LTD) the synapse (e.g., Cooper, Bloom, & Roth, 1991; Lynch et al., 1983; Wickens, 1993). Because the spines are somewhat separated from the bulk of the intracellular medium, free Ca^{2+} persists for several seconds after entering the cell (Gamble & Koch, 1987;

MacDermott et al., 1986). Under ideal conditions, the dopamine-mediated reward signal will arrive during this time, and there is substantial evidence that it will interact with the glutamate signal. The most popular model of this interaction assumes that after dopamine binds to the D_1 receptor and activates its associated G protein, a sequence of chemical reactions result that ultimately inhibit the deactivation of the Ca-dependent protein kinases that are activated after glutamate binds to the NMDA receptor (Nairn, Hemmings, Walaas, & Greengard, 1998; Pessin et al., 1994; Wickens, 1990, 1993). The effect of this inhibition is that dopamine locks the glutamate second messenger in the "on" position, thereby potentiating the learning effect. Thus, the presence of dopamine strengthens the synapses that were active on a trial when reward was delivered (e.g., Huang & Kandel, 1995).

The model described in Figures 5 and 6 easily accounts for all of the dissociations between rule-based and information integration category learning tasks that were described above. First, because the dopamine mediated reward signal is thought to be necessary for learning (e.g., LTP) to occur in the caudate nucleus, the absence of such a reward signal should greatly interfere with this form of implicit category learning. For this reason, the model predicts that learning in information integration tasks should be impaired (relative, say, to learning in rule-based tasks) during unsupervised categorization, or when the category label is shown before stimulus presentation (rather than after the response). In addition, as mentioned above, the timing of the reward signal relative to the response is critical for this type of learning. In reward-mediated learning, it is essential to strengthen those (and only those) synapses that actively participated in the response that elicited the reward. Because there is necessarily some delay between response and reward delivery, this means, therefore, that some trace must be maintained that signals which synapses were recently active. In the case of the medium spiny cells in the caudate nucleus, the morphology of the dendritic spines allows this trace to exist for several seconds after the response is initiated (Gamble & Koch, 1987; MacDermott et al., 1986). If the reward is delayed by more than this amount, then the ensuing dopamine release will strengthen inappropriate synapses and learning will be adversely affected.

The model described in this section does not make strong predictions about the effects of switching hands or response locations after learning is complete. This is because there are projections from the caudate nucleus to all frontal areas, including prefrontal, premotor, and motor cortices (via the globus pallidus and the thalamus; e.g. Alexander et al., 1986). Even so, the neostriatum (i.e., the caudate and putamen) has been strongly implicated in procedural motor learning (Jahanshahi et al., 1992; Mishkin et al., 1984; Saint-Cyr et al., 1988; Willingham et al., 1989), so it is not unexpected that an implicit category learning system situated in the tail of the caudate nucleus

would engage in response learning more strongly than, say, a rule-based system that is largely mediated within prefrontal cortex.

Finally, the model described here is also consistent with the dual-task study of Waldron and Ashby (2001). The numerical Stroop task that was used as the dual task in this study was selected specifically because it is known to activate frontal cortical areas. As such, it was predicted to interfere more strongly with the frontal-based explicit reasoning system than with the caudate-based implicit system.

In addition to accounting for these dissociations, the model described in Figures 5 and 6 also accounts for the dissociations that have been reported for various neuropsychological patient groups (i.e., summarized in Table 1). First, the model predicts category learning deficits in information-integration tasks in patients with Parkinson's or Huntington's disease because both of these populations suffer from caudate dysfunction. It also explains why frontal patients and medial temporal lobe amnesiacs are relatively normal in these tasks – that is, because neither prefrontal cortex nor medial temporal lobe structures play a prominent role in the Figure 5 model.

Before closing this section, it should be noted that the model shown in Figure 5 is strictly a model of *visual* category learning. However, it is feasible that a similar system exists in the other modalities, since they almost all also project directly to the basal ganglia, and then indirectly to frontal cortical areas (again via the globus pallidus and the thalamus; e.g., Chudler, Sugiyama, & Dong, 1995). The main difference is in where within the basal ganglia they initially project. For example, auditory cortex projects directly to the body of the caudate (i.e., rather than to the tail; Arnalud, Jeantet, Arsaut, & Demotes-Mainard, 1996).

6. A possible perceptual representation category learning system

No one has yet proposed a detailed category learning model that uses the perceptual representation memory system. However, based on work in the memory literature, it seems likely that such a category learning system, if it exists, would be based in sensory cortex (Curran & Schacter, 1996; Schacter, 1994) and would involve some form of perceptual learning. As mentioned above, it has been suggested that such a system might play a prominent role in prototype distortion tasks (Ashby & Ell, 2001).

Before investigating this possibility further, it is worth noting that even if the perceptual representation memory system did contribute to learning in prototype distortion tasks, it is not clear that prototype abstraction would meet the standard criteria of a separate system (Ashby & Ell, 2002b). When the stimuli are visual in nature, then any category learning system must receive input from the visual system. If some category learning system X depends on input from the brain region mediating prototype abstraction, then system X and the prototype abstraction system would not be mediated by separate neural pathways – a condition often considered necessary for separate systems (e.g., Ashby & Ell, 2002b). For example, under this scenario, a double dissociation between system X and the prototype system should be impossible. Damage to the neural structures downstream from visual cortex that mediate system X should induce deficits in category learning tasks mediated by system X, but not necessarily in prototype abstraction tasks. On the other hand, damage to visual cortex should impair all types of visual category learning. Thus, if prototype abstraction is mediated within visual cortex, then any group impaired in prototype abstraction should also be impaired on all other category learning tasks. In addition, it should be extremely difficult, or impossible, to find neuropsychological patient groups that are impaired in prototype abstraction, but not in other types of category learning. The available neuropsychological data supports this prediction, but as Table 1 indicates, only very limited tests of this prediction are currently possible.

Although the term "perceptual learning" is often broadly defined (e.g., Kellman, 2002), in this chapter we use the term to refer specifically to learning related changes in sensory cortex. Perceptual learning of this type is thought to occur any time repeated presentations of the same stimulus occur during some relatively brief time interval (Dosher & Lu, 1999). Unlike the reward-mediated learning that is thought to occur in the basal ganglia, no reward seems necessary for perceptual learning (e.g., Kellman, 2002). In fact, a response does not even seem to be required (e.g., Posner & Keele, 1968; Homa & Cultice, 1984). Presumably then, perceptual learning is mediated by a form of LTP that is quite close to classical Hebbian learning. In other words, rather than the three-factor learning rule described in the previous section in which learning occurs only in the presence of presynaptic activation, postsynaptic activation, and reward, apparently with perceptual learning, only pre- and postsynaptic activation are necessary.

In the visual cortex, LTP has been shown to occur at synapses between cortical pyramidal cells. Like the LTP that occurs in the procedural learning system, LTP in the perceptual representation system requires presynaptic

activation from cortical cells releasing glutamate. However, this system does not require activation of dopamine receptors for LTP to occur. In the procedural learning system, the activation of dopamine receptors eventually potentiates the learning-related effects of the protein kinases that are thought to be activated by the glutamate signal (Wickens, 1993). The most widely known mechanism of cortical LTP also requires activation of NMDA channels. As in medium spiny cells, activation of NMDA receptors in cortex leads to an increase in intracellular Ca^{2+}, and subsequently to an increase in a protein kinase that has been shown to mediate LTP (i.e., calcium dependent protein kinase II). Unlike medium spiny cells in the caudate nucleus, however, this process apparently does not require dopamine (i.e., the cortical protein kinase undergoes autophosphorylation) (Malenka & Nicoll, 1999).

Many different types of categorization experiments have been reported in the literature. Ashby and Maddox (1998) distinguished between what they called (A, B) tasks and (A, not A) tasks. In an (A, B) task, subjects are presented a series of exemplars that are each from some category A or from a contrasting category B. The task of the subject is to respond with the correct category label on each trial (i.e., "A" or "B"). In an (A, not A) task, there is a single central category A and subjects are presented with a series of stimuli that each are either an exemplar from category A or a stimulus that does not belong to category A. The subject's task is to respond "Yes" or "No" depending on whether the presented stimulus was or was not a member of category A. Historically, prototype distortion tasks have been run both in (A, B) form and in (A, not A) form. An important difference is that in an (A, B) task, the stimuli associated with both responses each have a coherent structure – that is, they each have a central prototypical member around which the other category members cluster (and likelihood tends to decrease monotonically with psychological distance from the prototype). In an (A, not A) task, this is true of the stimuli associated with the "A" (or "Yes") response, but not of the stimuli associated with the "not A" (or "No") response. The "not A" stimuli have no central member, no coherent structure, and over a reasonably large region of stimulus space, any given pattern is as likely to be associated with this response as any other pattern (with the exception of course, of the part of the space in which the category A exemplars are clustered).

This digression is important because if the perceptual representation system contributes to category learning, then it likely will have very different effects in (A, not A) and (A, B) tasks. Consider first an (A, not A) task. The category A prototype will induce a graded pattern of activation throughout visual cortex. One particular cell (or small group of cells) will fire most rapidly to the

presentation of this pattern. Call this cell A. In other words, cell A will fire to a particular range of visually similar patterns that includes the category A prototype. A low level distortion of the category A prototype will be visually similar to the prototype and therefore will also likely cause cell A to fire. Thus cell A will repeatedly fire throughout training on the category A exemplars. As a result, perceptual learning will cause the magnitude of the cell A response to increase throughout training. In contrast, the stimuli associated with the "not A" response will be visually dissimilar to the category A prototype and therefore will be unlikely to cause cell A to fire. During the transfer or testing phase of the experiment, the subject can use the increased sensitivity of cell A to respond accurately. In particular, stimuli from category A are likely to lead to an enhanced visual response compared to stimuli that do not belong to category A. From the subject's perspective, this enhanced visual response might be interpreted as an increased visual familiarity. Thus, to respond with above chance accuracy, subjects need only respond "A" or "Yes" to any stimulus that elicits a feeling of familiarity.

Next, consider an (A, B) task. In this case there will be some cell A maximally tuned to the category A prototype, but there will be some other cell B that is tuned to the category B prototype. During training, every presented stimulus is a distortion of either the category A or category B prototype, so it is likely that either cell A or B will fire on many trials. The actual number will depend on how much the prototypes are distorted to create the two categories. During the testing phase, all stimuli are again from either category A or B, and so stimuli from both categories will be equally likely to elicit an enhanced visual response (assuming the same level of distortion was used to create both categories). As a result, almost everything will feel familiar to the subject, so this feeling of familiarity will not help subjects decide whether to respond "A" or "B".

The conclusion therefore, is that if the perceptual representation system involves two-factor Hebbian learning, then that system could greatly assist in learning in (A, not A) tasks, but it would be of little help in (A, B) tasks. This is not to say that learning in (A, B) prototype distortion tasks is impossible, only that other learning systems must be used. Kolodny (1994) reported that amnesiacs learn normally in (A, B) prototype distortion tasks [actually in (A, B, C) tasks], so it seems unlikely that people memorize the category label associated with each prototype. One obvious possibility is that they instead use a procedural-memory based system of the type described in the previous section. If so, then several strong, yet untested predictions follow. First, patients with basal ganglia disease (e.g., Parkinson's or Huntington's disease) should be normal in (A, not A) prototype distortion tasks, but impaired in (A,

B) tasks. Second, because feedback is much more important to procedural learning than perceptual learning, unsupervised prototype distortion category learning should be better in (A, not A) tasks than in (A, B) tasks. Homa and Cultice (1984) showed that unsupervised learning is possible in (A, B) tasks if the category members are all low-level distortions of the prototypes, but to our knowledge, no one has systematically compared unsupervised learning in (A, not A) and (A, B) tasks.

Because there is so little available data, the predictions and inferences drawn in this section are highly speculative. Therefore, much more work needs to be done before we will have a clear understanding of the role played by the perceptual representation memory system in category learning.

7. Summary and conclusions

The issue of whether human category learning is mediated by one or several category learning systems is a question of intense current debate. Although this issue is still unresolved, recent cognitive, neuropsychological, and neuroimaging data support the weaker hypothesis that different memory systems may participate in different types of category learning tasks. This chapter focused on two memory systems that may contribute to implicit category learning – procedural memory and the perceptual representation memory system.

The recent surge of interest in implicit category learning has a number of practical benefits. First, it immediately ties the categorization literature to the large and well established memory literature. Second, it organizes new research efforts, and it encourages collecting data of a qualitatively different nature than have been collected in the past. Third, it encourages a more critical examination of categorization theories than has been common in the past – largely because it adds constraints on both psychological process and neural structure that historically have not received much attention in the categorization literature. Thus, no matter how it is eventually resolved, the field will benefit from the current interest in implicit categorization.

Notes

* This research was supported in part by National Science Foundation Grant BCS99-75037. We thank Luis Jimenez and Eliot Hazeltine for their helpful comments. Correspondence concerning this chapter should be addressed to F. Gregory Ashby, Department of Psychology, University of California, Santa Barbara, CA 93106 (e-mail: ashby@psych.ucsb.edu).

1. Crick and Koch (1998) did not take the strong position that working memory is necessary for conscious awareness. Even so, they did argue that some short-term memory store is required. However, they left open the possibility that an extremely transient iconic memory might be sufficient.

2. Note that there is an explicit rule that also yields perfect accuracy, but it involves three "ands" and two "ors". Despite running many subjects through the Figure 2 categories, we have never had a subject describe this explicit rule at the end of training, even though almost all subjects eventually learn these categories perfectly. For this reason, the Figure 2 task is better described as an information-integration task, rather than as a rule-based task.

References

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, *9*, 357-81.

Alfonso-Reese, L. A. (1997). Dynamics of category learning. Unpublished doctoral dissertation, University of California, Santa Barbara.

Arnalud, E., Jeantet, Y., Arsaut, J., & Demotes-Mainard, J. (1996). Involvement of the caudal striatum in auditory processing: c-fos response to cortical application of picrotoxin and to auditory stimulation. *Brain Research: Molecular Brain Research, 41*, 27-35.

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*, 442-81.

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

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

Ashby, F. G., & Ell, S. W. (2002b). Single versus multiple systems of learning and memory. In J. Wixted & H. Pashler (Eds.), *Stevens' handbook of experimental psychology: Vol. 4 Methodology in experimental psychology* (3rd ed., pp. 655-691). New York: Wiley.

Ashby, F. G., Ell, S. W., & Waldron, E. M. (2002). Abstract category labels are learned in rule-based categorization, but response positions are learned in information-integration categorization. Manuscript under review.

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

Ashby, F. G. & Maddox, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Handbook of perception & cognition: Judgment, decision making, and measurement* (Vol. 3). New York: Academic Press.

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

Ashby, F. G., Noble, S., Filoteo, J. V., Waldron, E. M., & Ell, S. W. (2002). Category learning deficits in Parkinson's disease. Manuscript submitted for publication.

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

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

Ashby, F. G., Waldron, E. M., Lee, W. W., & Berkman, A. (2001). Suboptimality in human categorization and identification. *Journal of Experimental Psychology: General*, *130,* 77-96.

Atalay, J. & Wise, R. A. (1983). Time course of pimozide effects on brain stimulation reward. *Pharmacology, Biochemistry and Behavior, 18,* 655-658.

Beninger, R. J. (1983). The role of dopamine in locomotor activity and learning. *Brain Research, 287,* 173-196.

Bozarth, M. A. (1994). Opiate reinforcement processes: Re-assembling multiple mechanisms. *Addiction, 89,* 1425-1434.

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

Brown, R. G. & Marsden, C. D. (1988). Internal versus external cures and the control of attention in Parkinson's disease. *Brain, 111*, 323-345.

Bruner, J. S., Goodnow, J., & Austin, G. (1956). *A study of thinking.* New York: Wiley.

Chudler, E. H., Sugiyama, K., & Dong, W. K. (1995). Multisensory convergence and integration in the neostriatum and globus pallidus of the rat. *Brain Research, 674*, 33-45.

Cools, A.R., van den Bercken, J.H L., Horstink, M.W.I., van Spaendonck, K.P.M., & Berger, H.J.C. (1984). Cognitive and motor shifting aptitude disorder in Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 47*, 443-453.

Cooper, J. R., Bloom, F. E., & Roth, R. H. (1991). *The biochemical basis of neuropharmacology (Sixth Edition)*. New York: Oxford.

Crick, F. & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in Neuroscience, 2*, 2263-2275.

Crick, F. & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature, 375*, 121-123.

Crick, F. & Koch, C. (1998). Cousciousness and neuroscience. *Cerebral Cortex, 8*, 97-107.

Curran, T. & Schacter, D. L. (1996). Memory: Cognitive neuropsychological aspects. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioral Neurology and Neuropsychology* (pp. 463-471). New York: McGraw-Hill.

Difiglia, M., Pasik, T., & Pasik, P. (1978). A Golgi study of afferent fibers in the

neostriatum of monkeys. *Brain Research, 152,* 341-347.
Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research*, *39*, 3197-3221.
Eacott, M. J., & Gaffan, D. (1991). The role of monkey inferior parietal cortex in visual discrimination of identity and orientation of shapes. *Behavioural Brain Research, 46,* 95-98.
Elliott, R. & Dolan, R. J. (1998). Activation of different anterior cingulate foci in association with hypothesis testing and response selection. *Neuroimage, 8*, 17-29.
Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*, 107-140.
Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, *18*, 500-549.
Estes, W. K. (1994). *Classification and cognition*. Oxford: Oxford University Press.
Filoteo, J. V., Maddox, W. T., & Davis, J. (2001a). A possible role of the striatum in linear and nonlinear categorization rule learning: Evidence from patients with Huntington's disease. *Behavioral Neuroscience*, *115*, 786-798.
Filoteo, J. V., Maddox, W. T., & Davis, J. D. (2001b). Quantitative modeling of category learning in amnesic patients. *Journal of the International Neuropsychological Society*, *7*, 1-19.
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 spine. *Neuroscience, 13,* 1189-1215.
Fuster, J. M. (1989). *The prefrontal cortex* (2nd Edition). New York: Raven Press.
Gaffan, D. & Eacott, M. J. (1995) Visual learning for an auditory secondary reinforcer by macaques is intact after uncinate fascicle section: indirect evidence for the involvement of the corpus striatum. *European Journal of Neuroscience*, *7,* 1866-1871.
Gaffan, D. & Harrison, S. (1987). Amygdalectomy and disconnection in visual learning for auditory secondary reinforcement by monkeys. *Journal of Neuroscience, 7,* 2285-2292.
Gamble, E. & Koch, C. (1987). The dynamics of free calcium in dendritic spines in response to repetitive synaptic input. *Science, 236,* 1311-1315.
Gluck, M. A., Oliver, L. M., & Myers, C. E. (1996). Late-training amnesic deficits in probabilistic category learning: A neurocomputational analysis. *Learning and Memory, 3*, 326-340.
Goldman-Rakic, P. S. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational knowledge. in *Handbook of Physiology* (Plum, F. & Mountcastle, V., eds.), pp. 373-417, American Physiological Society.
Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron, 14*, 477-485.
Heaton, R. K. (1981). A manual for the Wisconsin Card Sorting Test. Odessa, FL: Psychological Assessment Resources.
Homa, D. & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 83-94.
Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based

generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 418-439.

Huang, Y. Y. & Kandel, E. R. (1995). D1/D5 receptor agonists induce a protein synthesis-dependent late potentiation in the CA1 region of the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America, 92,* 2446-2450.

Jahanshahi, M., Brown, R. G., & Marsden, C. (1992). The effect of withdrawal of dopaminergic medication on simple and choice reaction time and the use of advance information in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *55*, 1168-1176.

Janowsky, J. S., Kritchevsky, A. P., & Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral Neuroscience, 103,* 548-560.

Kellman, P. J. (2002). Perceptual learning. In R. Gallistel & H. Pashler (Eds.), *Stevens' handbook of experimental psychology: Vol. 3 Learning, motivation, and emotion* (3rd ed., pp. 259-299). New York: Wiley.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399-1402.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, *3*, 172-179.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747-1749.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory*, *1*, 106-120.

Knowlton, B. J., Squire, L. R., Paulsen, J. S., Swerdlow, N. R., Swenson, M., & Butters, N. (1996). Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology*, *10*, 538-548.

Köhler, W. (1925). *The mentality of apes*. New York: Harcourt, Brace & Co.

Kolb, B., & Whishaw, I. Q. (1990). *Fundamentals of Human Neuropsychology* (3rd Ed.). New York: W. H. Freeman & Company.

Kolodny, J. A. (1994). Memory processes in classification learning: An investigation of amnesic performance in categorization of dot patterns and artistic styles. *Psychological Science, 5*, 164-169.

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*, 359-362.

Lynch, G., Larson, J., Kelso, S., Barrionuevo, G., & Schottler, F. (1983). Intracellular injections of EGTA block induction of hippocampal long-term potentiation. *Nature, 305,* 719-721.

MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J., & Barker, J. L. (1986). NMDA-receptor activation increases cytoplasmic calcium concentration in cultured spinal cord neurones. *Nature, 321,* 519-522.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar

models of categorization. *Perception and Psychophysics, 53*, 49-70.

Maddox, W. T. & Ashby, F. G., & Bohil, C. J. (2002). Delayed feedback effects on rule-based and information-integration category learning. Manuscript under review.

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

Malenka, R. C., & Nicoll, R. A. (1999). Long-term potentiation--a decade of progress? *Science, 285,* 1870-1874.

McDonald, R. J. & White, N. M. (1993). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience, 107*, 3-22.

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

Medin, D. L. & Schaffer, M. M. (1978) Context theory of classification learning. *Psychological Review, 85*, 207-238.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1997). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*, 242-279.

Meulemans, T., Peigneux, P., & Van der Linden, M. (1998). Preserved artificial grammar learning in Parkinson's disease. *Brain & Cognition, 37,* 109-112.

Miller, J. D, Sanghera, M. K., & German, D. C. (1981). Mesencephalic dopaminergic unit activity in the behaviorally conditioned rat. *Life Sciences, 29,* 1255-1263.

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 3,* 775-799.

Mishkin, M., Malamut, B., & Bachevalier, J. (1984). Memories and habits: Two neural systems. In G. Lynch, J. L. McGaugh, & N. M. Weinberger (Eds.), *Neurobiology of human learning and memory* (pp. 65-77). New York: Guilford.

Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin card sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *21*, 7733-7741.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16,* 1936-1947.

Murphy, G. L. & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Nairn, A. C., Hemmings, H. C., Walaas, S. I., & Greengard, P. (1988). DARPP-32 and phosphatase inhibitor-1, two structurally related inhibitors of protein phosphatase-1, are both present in striatonigral neurons. *Journal of Neurochemistry, 50,* 257-262.

Nosofsky, R. M. (1986) Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*,

7, 375-402.

Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review, 9*, 169-174.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 247-255.

Packard, M. G., Hirsh, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *Journal of Neuroscience*, *9*, 1465-1472.

Packard, M. G. & McGaugh, J. L. (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: Further evidence for multiple memory systems. *Behavioral Neuroscience*, *106*, 439-446.

Pessin, M. S., Snyder, G. L., Halpain, S., Giraut, J.-A., Aperia, A., & Greengard, P. (1994). DARPP-32/protein phosphatase-1/Na+/K+ ATPase system: A mechanism for bidirectional control of cell function. In K. Fuxe, L F. Agnat, B. Bjelke, & D. Ottoson (Eds.), *Trophic regulation of the basal ganglia* (pp. 43-57). New York: Elsevier Science.

Pfaus, J. G. & Phillips, A. G. (1991). Role of dopamine in anticipatory and consummatory aspects of sexual behavior in the male rat. *Behavioral Neuroscience, 105,* 727-743.

Phillips, A. G., Blaha, C. D., Pfaus, J. G., & Blackburn, J. R. (1992). Neurobiological correlates of positive emotional states: Dopamine, anticipation and reward. In: Ken T. Strongman, Ed. *International review of studies on emotion*, Vol. 2.. New York, NY: John Wiley & Sons.

Pickering, A. D. (1997). New approaches to the study of amnesic patients: What can a neurofunctional philosophy and neural network methods offer? *Memory, 5*, 255-300.

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

Posner, M. I. & Petersen, S. E. (1990). Attention systems in the human brain. *Annual Review of Neurosci*ence*, 13*, 25-42.

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, 27,* 1987-1993.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *18*, 219-235.

Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning and Memory, 1*, 217-229.

Reber, P J., Stark, C. E. L., and Squire, L. R. (1998). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory, 5*, 420-428.

Robinson, A. L., Heaton, R. K., Lehman, R. A. W., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions.

Journal of Consulting and Clinical Psychology, 48, 605-614.

Rolls, E. T. (1994). Neurophysiology and cognitive functions of the striatum. *Revue Neurologique, 150,* 648-660.

Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1988). Procedural learning and neostriatal dysfunction in man. *Brain*, *111*, 941-959.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501-518.

Schacter, D. L. (1994) Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In D. L. Schacter & E. Tulving (Eds.), *Memory Systems 1994* (pp. 233-268). Cambridge: MIT Press.

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, 9,* 5720-5724.

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

Smith, E. E. & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Waldron, E. M. & Ashby, F. G. (2001). The effects of concurrent task interference on category learning. *Psychonomic Bulletin & Review*, *8*, 168-176.

White, N. M. (1989). A functional hypothesis concerning the striatal matrix and patches: mediation of S-R memory and reward. *Life Sciences, 45,* 1943-1957.

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.

Wickens, J. (1993). *A theory of the striatum*. New York: Pergamon Press.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1047-1060.

Wilson, C. J. (1995). The contribution of cortical neurons to the firing pattern of striatal spiny neurons. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia* (pp. 29-50). Cambridge: Bradford.

Winocur, G. & Eskes, G. (1998). Prefrontal cortex and caudate nucleus in conditional associative learning: Dissociated effects of selective brain lesions in rats. *Behavioral Neuroscience, 112*, 89-101.