Chapter 5

The neurobiology of categorization

F. Gregory Ashby and Matthew J. Crossley

Editors' Preview

Chapter 5, 'The neurobiology of categorization', provides a review of what is known about the neural bases for categorization. A major theme of the chapter is that categories characterized by different kinds of structures may enlist different types of learning and memory processes that in turn rely on different brain circuitry. For example, as noted in Chapters 2, 3, and 4, human adults possess the ability to rapidly learn rule-based categories in which the rule splitting the categories is explicit, verbalizable, accessible to conscious awareness (e.g. red things go into group *A*, green things go into group *B*), and believed to be stored in a declarative form of memory that enables recall of facts and events.

Human adults also learn categories that require an integration of information that is not readily verbalizable, may be slowly acquired, and may be stored in a nondeclarative procedural memory system. Figure 5.1 provides an example of categories that can be learned through an integration of information about the orientation and thickness of the bars within the circular discs on each side of the boundary. Examples of real-world categories learned through processes of information integration include determining whether X-rays provide evidence of cancerous tumours and (as was referenced in Chapter 2) determining whether day-old chicks are male or female.

Various lines of evidence acquired with different investigative techniques within cognitive neuroscience suggest that categories acquired through explicit rule learning versus information integration may be processed through two distinct brain pathways. The evidence includes studies of single-unit recording in animals (an electrode is used to register the activity of an individual neuron as an animal is performing a categorization task), brain stimulation (a brain region is directly stimulated with electric current as an animal learns a categorization task), lesion studies (a brain region of an animal is damaged to determine what effect that has on categorization performance),

functional magnetic resonance imaging or fMRI (allowing one to determine which parts of the brain are active during a categorization task by tracking which brain regions are using the most oxygen), and individuals who have suffered brain damage (either through stroke, injury, or disease processes, to determine what effect the damage has on categorization abilities).

Evidence suggests that highly practised category judgements are mediated by a third brain pathway that is qualitatively distinct from both the rule-based and information-integration brain pathways that dominate early category learning. This notion is consistent with the idea that over time and with practice, categories that are initially slowly acquired can become more rapidly accessed in an automatic way. More plainly, the idea is that one can distinguish at both cognitive and neural levels between processes of category formation on the one hand, and category possession on the other hand.

The suggested scheme for the neurobiology of categorization has both phylogenetic and ontogenetic implications that are relevant to the major themes of the book. As is observed by the authors of this chapter, comparatively, one can examine particular differences in brain structures for humans and nonhuman animals that are believed to mediate categorization performance to make specific behavioural predictions for how performance should vary across species. In addition, developmentally, one can examine how early in infancy or childhood the two systems become functional and whether one system might emerge subsequently to the other.

Categorization is the act of responding differently to objects or events in separate classes or categories. It is a vitally important skill that allows us to approach friends and escape foes, to find food, and avoid toxins. The scientific study of categorization has a long history. For most of this time, the focus was on the cognitive processes that mediate categorization. Within the past decade, however, considerable attention has shifted to the study of the neural basis of categorization. This chapter reviews that work.

There is recent evidence that highly practised categorization judgements are mediated by neural pathways different from those that mediate initial category learning; so our review discusses category learning and automatic categorization judgements separately. We begin in the next section with a brief overview of the basal ganglia, which are a collection of subcortical nuclei that are especially important in categorization. Next, we focus on initial category learning. We then consider the neural basis of automatic categorization judgements. Finally, we close with some general comments and conclusions.

The basal ganglia

Many brain areas are thought to participate in category learning, including the prefrontal cortex (PFC), the hippocampus and other medial temporal lobe structures, the anterior cingulate, and the basal ganglia. The basal ganglia seem especially important. For example, they are one of the few areas that have been identified in virtually every cognitive or behavioural neuroscience study of category learning (for a review, see Ashby & Ennis, 2006). As they play such a prominent role in the cognitive neuroscience literature on categorization, this section briefly reviews basal ganglia neuroanatomy and some key evidence implicating the striatum in category learning.

The basal ganglia are a collection of subcortical brain regions that can be divided into input structures, output structures, and collections of cells that produce the neurotransmitter dopamine. The input structures include the caudate nucleus and putamen (and other regions), which together are often referred to as the dorsal striatum or just the striatum. The striatum receives inputs from almost all of the cortex, and these cortical–striatal projections are characterized by massive convergence (Kincaid et al., 1998). Roughly speaking, the putamen receives input from parietal and motor areas, and the caudate receives input from visual and auditory association areas and from frontal cortex (e.g. Heimer, 1995). The anterior region of the caudate, often called the head of the caudate, is reciprocally connected to the PFC and so is often implicated in executive function tasks. The posterior (or caudal) caudate, often called the tail of the caudate, receives input from the visual association cortex (the body of the caudate is the region between the head and tail). Thus, the striatum is in a unique position in the mammalian brain since it receives direct, but highly compressed input from virtually the entire cortex.

Compared to cortex, the structure of the striatum is extremely simple. It contains a single layer composed predominantly of medium spiny cells. The dendrites of these cells receive input from the axons of around 15 000 cortical pyramidal cells and the medium spiny cell axons project out of the striatum to basal ganglia output structures, such as the globus pallidus. The medium spiny cells are GABAergic, and hence inhibitory, with a low spontaneous firing rate. The cells in the globus pallidus, which are also GABAergic, project to the thalamus, which in turn, sends excitatory projections to cortex. Spontaneous activity in the globus pallidus is high (e.g. Wilson, 1995), so the globus pallidus tonically inhibits the thalamus. Cortical activation of the striatum however, causes the medium spiny cells to inhibit the pallidal cells, thereby releasing the thalamus from its tonic inhibition. Due to this functional architecture, the basal ganglia are frequently described as applying a brake on cortex because they tonically prevent the thalamus from stimulating cortex. Cortex can release the brake by stimulating the striatum.

Another feature of the striatum that likely contributes to the important role that it plays in category learning is that cortical–striatal synapses display an unusual form of synaptic plasticity that is ideally suited to feedback-dependent learning. In fact, the conditions required to change the efficacy of cortical–striatal synapses (i.e. to induce long-term potentiation and long-term depression) closely match the conditions for reinforcement learning (Sutton & Barto, 1998). Specifically, the best available evidence

suggests that cortical–striatal synapses that were active on rewarded trials are strengthened and synapses that were active on trials that do not produce reward are weakened. The key is the neurotransmitter dopamine, which is released into the striatum from the substantia nigra pars compacta. Many studies have shown that dopamine release increases above baseline following unexpected reward, and it falls below baseline following unexpected absence of reward (e.g. Schultz et al., 1997). Many researchers have proposed that dopamine serves as the reinforcement signal in striatal-based reinforcement learning (e.g. Ashby et al., 2007; Houk et al., 1995; Wickens, 1993).

The earliest proposals that the striatum may be an important locus of category learning (Ashby et al., 1998) were motivated by a long series of lesion studies in rats and monkeys supporting the hypothesis that the striatum is both necessary *and* sufficient for visual discrimination learning. Technically, a discrimination task is a special case of categorization in which each category contains only one exemplar.

Many studies have shown that lesions of the striatum 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; Packard & McGaugh, 1992). For example, in one study, rats with striatal lesions 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 signalling 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 striatum (e.g. Ashby et al., 2007; Wickens, 1993).

More recent single-unit recording studies confirm the striatal contribution to categorization. In a series of studies, monkeys were taught to classify a tactile stimulus (i.e. a rod dragged across the monkey's finger) as either 'low speed' or 'high speed' (e.g. Merchant et al., 1997). A large number of cells in the putamen showed learning-related changes in their firing properties. For example, after training, many cells fired to any movement in the low-speed category, but not to movements in the high-speed category (or vice versa). These same neurons were not active during passive experience with the stimuli or during a control motor task. Furthermore, the activity of these neurons predicted the behaviour of the monkeys.

There are several prominent basal ganglia diseases that have played an important role in attempts to understand the role of the basal ganglia in categorization. The most theoretically important of these are Parkinson's disease and Huntington's disease. In Parkinson's disease, the dopamine cells die and this depletes the striatum of the key neurotransmitter that mediates cortical–striatal plasticity. In Huntington's disease, the medium spiny cells in the striatum die.

The neurobiology of category learning

There is now overwhelming evidence that humans have multiple category-learning systems. By definition, learning requires some change in the brain that persists beyond

the training event. If this change is defined as a memory trace, then it seems plausible that all major memory systems should be capable of some form of category learning (Ashby & O'Brien, 2005). Different memory systems have different properties, and each is ideally suited to a certain type of learning. In fact, we review considerable evidence that the different systems are ideally suited to learning about different types of category structures. There is evidence that changing the category structure can induce a switch from one system to another, and as a result, can cause the resulting data to differ in a qualitative manner. When reviewing the category-learning literature, it is therefore critical that we pay particular attention to which memory system is best suited to learning the type of category structures used in the research.

Declarative memory systems

Declarative memories are those accessible to conscious awareness (Eichenbaum, 1997). This includes working memory and episodic/semantic memory. Working memory is the ability to maintain and manipulate limited amounts of information during brief periods of cognitive activity (Baddeley, 1986). It is heavily used in explicit reasoning and problem-solving tasks, and more generally in any task said to depend on executive function. A huge literature is devoted to the neural basis of working memory (see, e.g. Fuster, 1989). Suffice it to say that there is overwhelming evidence linking working memory to the PFC. However, there is also evidence that the head of the caudate nucleus is critical in working-memory tasks (Hikosaka et al., 1989; Schultz & Romo, 1992), and a variety of models have been proposed that assume working memory is mediated, at least in part, by cortical–striatal loops (e.g. Ashby et al., 2006; Frank et al., 2001; Monchi et al., 2000).

Episodic and semantic memories are known to depend on the hippocampus and other medial temporal lobe structures (e.g. Squire & Schacter, 2002; Tulving, 2002). For example, medial temporal lobe damage often causes anterograde amnesia – that is, deficits on tests that depend on recent declarative memories.

As working memory is effective only for short time periods, it cannot store long-term category knowledge, but it could be the primary learning system in tasks where the categories are learned quickly. However, even in this case, episodic and semantic memory systems would presumably be required to consolidate this learning, so one would likely expect the different declarative systems to work together during category learning.

Declarative memory systems might mediate learning in any categorization task in which subjects are consciously aware of their categorization strategy. Obvious possibilities include tasks in which subjects formulate and test explicit hypotheses about category membership or explicitly memorize the category labels associated with each stimulus.

Tasks that encourage explicit hypothesis testing

In *rule-based category-learning tasks,* the categories can be learned via an explicit hypothesis-testing procedure (although, of course, they could also be learned in other ways). In most cases the rule that maximizes accuracy is easy to describe verbally (Ashby et al., 1998). In the simplest examples, only one stimulus dimension is relevant, and the subject's task is to discover this dimension and then to map the different

dimensional values to the relevant categories. More difficult rule-based tasks require attention to two or more dimensions. For example, the correct rule might be a conjunction of the type: 'the stimulus is in category *A* if it is large and bright'. The key requirement is that the correct categorization rule in rule-based tasks is one that can be discovered by an explicit hypothesis-testing procedure. Virtually all category-learning tasks used in neuropsychological assessment are rule based, including the widely known Wisconsin card sorting test (WCST) (Heaton, 1981). Stimuli in this task are cards containing geometric patterns that vary in colour, shape, and symbol number, and in all cases the correct categorization rule is one dimensional and easy to describe verbally.

Many researchers have proposed that people generate and test explicit hypotheses about category membership in rule-based tasks (e.g. Ashby et al., 1998; Erickson & Kruschke, 1998; Nosofsky et al., 1994), and Ashby and O'Brien (2005) argued that rule-based learning is mediated primarily by declarative memory systems. The idea is that working memory is used to store hypotheses about category membership during their testing, and episodic/semantic memory systems are used for the long-term storage and consolidation of these rules. A wide variety of evidence supports this hypothesis. For example, several studies have reported that a dual task requiring working memory and executive attention has massive detrimental effects on rule-based category learning, but little or no effect on the ability of subjects to learn other types of category structures (i.e. information-integration categories; DeCaro et al., 2008; Waldron & Ashby, 2001; Zeithamova & Maddox, 2006).

The only neurobiologically detailed model of category learning, called COVIS (Competition between Verbal and Implicit Systems), assumes that humans have separate hypothesis-testing and procedural-learning systems (Ashby et al., 1998; Ashby & Valentin, 2005; Ashby & Waldron, 1999). COVIS assumes that learning in rule-based tasks is dominated by an explicit, hypothesis-testing system that uses working memory and executive attention and is mediated primarily by the anterior cingulate, the PFC, the hippocampus, and the head of the caudate nucleus. There are two main sub-networks in this model: one that generates or selects new candidate hypotheses, and one that maintains candidate rules in working memory during the testing process and mediates the switch from one rule to another. The COVIS hypothesis-testing system is similar to the neural network models of the WCST that were proposed by Monchi et al. (2001) and Amos (2000).

COVIS and the cognitive neuroscience literature on working memory predict that the PFC and the head of the caudate nucleus should be active in rule-based tasks. A wide variety of evidence supports this prediction. First, impaired performance on the WCST is among the most classic of all signs of PFC damage (e.g. Kimberg et al., 1997). Second, many animal-lesion studies have confirmed the important role played by the PFC in rule learning and use (for a review, see Roberts & Wallis, 2000). For example, Joel et al. (1997) reported that lesions to the medial PFC in rats impaired the animals on a simplified version of the WCST. Third, a number of neuroimaging studies have used the WCST or a rule-based task similar to the WCST, and all of these have reported task-related activation in the PFC (e.g. Konishi et al., 1999; Monchi et al., 2001; Rogers et al., 2000).

There is also strong evidence that the striatum is critical to rule-based category learning. First, there are many reports that Parkinson's disease patients are impaired in rule-based tasks (e.g. Ashby et al., 2003; Brown & Marsden, 1988; Cools et al., 1984). Although later in the disease Parkinson's patients have frontal damage (primarily the result of cell death in the ventral tegmental area), the disease mainly targets the basal ganglia. The caudate region most affected appears to be the head of the caudate nucleus (van Domburg & ten Donkelaar, 1991), which is reciprocally connected to the PFC. Second, a number of functional neuroimaging studies of the WCST have reported task-related activation in the head of the caudate nucleus (e.g. Monchi et al., 2001; Rogers et al., 2000).

Explicit hypothesis testing requires working memory, but it also requires a number of other executive processes, including executive attention, rule selection, rule representation, and rule switching. Many studies have focused on the neural basis of one or more these processes.

There is a large literature on executive attention, which we do not review here. However, one relevant proposal is that executive attention is mediated in part by the PFC and anterior cingulate (Posner & Petersen, 1990). The (dorsal) anterior cingulate was also identified as a site of hypothesis generation in rule-based category learning (Elliott et al., 1999).

Important evidence about the neural representation of rules was obtained in a series of single-unit recording studies in which monkeys were taught to classify objects by applying either one rule (e.g. spatial) or another (e.g. associative). Each trial began with a cue signalling the animal about which rule to use with the ensuing stimulus. Several studies using this paradigm reported many PFC cells that showed rule-specific activity – that is, they fired during application of one of the rules (but not during the other), regardless of which stimulus was shown (Asaad et al., 2000; Hoshi et al., 1998; White & Wise, 1999).

A separate line of research implicates the head of the caudate nucleus in rule switching. First, stimulation of the striatum increases switching from one motor activity to another when such switching behaviours are rewarded (Jaspers et al., 1990). Second, lesioning the dopamine projection into the PFC *improves* the performance of monkeys in an analogue of the WCST, even though it impairs their spatial working memory (Roberts et al., 1994). This result at first seems perplexing, but it turns out that such lesions increase dopamine levels in the basal ganglia (Roberts et al., 1994). Therefore, if the basal ganglia are responsible for switching, and if switching is enhanced by dopamine, then lesioning dopamine fibres into the PFC should improve switching, which is exactly what Roberts et al. (1994) found. Third, van Golf Racht-Delatour and El Massioui (1999) reported that rats with lesions to the striatum had no deficits in learning which arm of a radial arm maze was initially baited, but they did have deficits when the position of the baited arm was successively switched according to a simple rule. Fourth, human fMRI studies have reported striatal activation during rule switching (e.g. Crone et al., 2005).

In summary, there is strong evidence that the PFC and head of the caudate nucleus are critical in rule-based tasks, and there is also evidence that the anterior cingulate may participate via its role in executive attention and hypothesis generation. On the

other hand, several studies have reported that medial temporal lobe amnesiacs are normal in rule-based category learning (Janowsky et al. 1989; Leng & Parkin, 1988). An obvious possibility is that many rule-based tasks are simple enough (e.g. the WCST) and that working memory is sufficient for subjects to keep track of which alternative rules they have tested and rejected. If so, then a natural prediction is that medial-temporal-lobe amnesiacs should be impaired in complex rule-based tasks (e.g. when the optimal rule is disjunctive). To our knowledge, this prediction has not been tested.

Tasks that encourage explicit memorization

In *unstructured category-learning tasks*, the exemplars of each category lack any coherent structure that could be discovered via hypothesis testing. Typically, unstructured categories are created by randomly assigning a set of perceptually distinct stimuli to the contrasting categories. For example, a category such as 'my personal numbers' (e.g. phone numbers, zip code, social security number) is unstructured because there is no logical rule or similarity relationship binding these numbers together. Introspection seems to suggest that such categories can only be learned via explicit memorization. Unstructured categories are closely related to ad hoc categories in the cognitive literature (Barsalou, 1983; see also Chapter 1 in this book) and pseudocategories in the animal literature (Lea, 1984).

Historically, tasks using unstructured categories have received relatively little attention. Most efforts have been simply to show that unstructured categories are among the most difficult of all types of category structures to learn (Nosofsky et al., 1994; Shepard et al., 1961; Smith et al., 2004).

In the only known neuroimaging study using unstructured categories, successful categorization was associated with activation in the body and tail of the caudate nucleus and the putamen, but not in the medial temporal lobes (Seger & Cincotta, 2005). On the neuropsychological front, Bayley et al. (2005) reported that two patients with dense amnesia, both as a result of bilateral medial temporal lobe lesions, were nonetheless able to learn difficult unstructured categories over the course of several weeks of training, despite not being able to describe the task, the instructions, or the stimuli at the start of each session.

Introspection seems to suggest that the only way arbitrary categories of this type could be learned is via explicit memorization. Based largely on this intuition, Ashby and O'Brien (2005) speculated that the medial temporal lobes would be critical for learning unstructured categories. The Seger and Cincotta (2005) results and the Bayley et al. (2005) results, however, cast serious doubt on this hypothesis. An alternative seems to be that the medial temporal lobes might mediate our conscious recollection of specific stimuli, but that the striatum is required to associate a category response with each stimulus because of the trial-by-trial reward-mediated learning that is required. Clearly, much more work is needed on the neural basis of unstructured category learning.

What role for the hippocampus?

The failure to find a more significant role for the hippocampus in rule-based and unstructured category-learning tasks is surprising for several reasons. First, the

hippocampus and other medial temporal lobe structures play a prominent role in the more traditional memory literature. Therefore, if category learning is a process of encoding and consolidating memories about categories, then it is natural to expect that the hippocampus might play a similar prominent role in category learning. Second, as described in other chapters in this book (e.g. Chapter 4), exemplar theory has been the most prominent cognitive theory of categorization for 30 years. Although there currently is no detailed neurobiological interpretation of exemplar theory, initial attempts to ground exemplar models of categorization in neurobiology have all assigned a key role to the hippocampus (e.g. Pickering, 1997; Sakamoto & Love, 2004). Thus, the success of exemplar theory also directs attention to the hippocampus.

To begin, it is important to note that the mapping of exemplar theory to the hippocampus is problematic. This is partly because the memory system postulated by exemplar theory appears qualitatively different from all of the memory systems that have been identified by memory researchers. According to exemplar theory, the memory representations that must be accessed prior to each categorization response are of each previously seen exemplar. These memory representations are detailed replicas of each exemplar (filtered by attentional processes), but they do not typically include contextual information (e.g. details about the experimental room). Thus, the closest match in the memory literature is probably semantic memory.

One problem with assuming that exemplar memory is a form of semantic memory is that semantic memory is declarative. In contrast, exemplar theorists are careful to assume that people do *not* have conscious awareness of exemplar memories. Thus, exemplar theory appears to postulate a unique memory system that has not yet been discovered by memory researchers.

It is important to note, however, that other instance-based theories postulate more traditional memory systems. For example, RULEX (Nosofsky et al., 1994) assumes people use explicit rules during categorization but they memorize exceptions. Presumably, people are aware of these exceptions, so this form of memory seems identical to semantic memory.

An obvious prediction of instance- or exemplar-based categorization theories, therefore, seems to be that patients with damage to medial temporal lobe structures should be impaired in category learning. We know of two studies that have reported category-learning deficits in amnesiacs (Kolodny, 1994; Zaki et al., 2003), and two others that reported normal performance on the first 50 trials, but impaired performance later on (Hopkins et al., 2004; Knowlton et al., 1994). On the other hand, many more studies have reported normal category-learning performance in patients with amnesia. First, many studies have shown that amnesiacs are normal in rule-based tasks such as the WCST (e.g. Janowsky et al., 1989; Leng & Parkin, 1988). Second, Filoteo et al. (2001b) reported normal performance by amnesiacs in a difficult information-integration task with nonlinearly separable categories that required hundreds of training trials. In fact, in the Filoteo et al. (2001b) study, one (medial temporal lobe) amnesiac and one control subject completed a second day of testing. Despite lacking an explicit memory of the previous session, the patient with amnesia performed slightly better than the control on the first block of day 2. This result suggests that amnesiacs do not necessarily rely on working memory to perform normally in category-learning tasks

(because working memory cannot be used to retain category knowledge across days). Third, several studies have also reported that patients with amnesia have normal performance in the (*A*, not *A*) prototype-distortion task that we describe in a later section (Knowlton & Squire, 1993; Kolodny, 1994; Squire & Knowlton, 1995; Zaki et al., 2003).

In summary, there currently is no neurobiological interpretation of exemplar theory. It seems likely that such an interpretation would have to assign a key role to medial temporal lobe structures. Some evidence suggests that medial temporal lobe damage can cause category-learning deficits under certain special conditions, but many studies have reported normal category learning in patients with amnesia. The role of the medial temporal lobes in categorization is likely to be a topic of intense research in future years.

The procedural memory system

Procedural memories are nondeclarative memories of skills that are learned through practice (Willingham, 1998). Traditionally, these have been motor skills, such as those used when playing the piano. Procedural learning is qualitatively different from learning that is mediated by declarative memory systems. First, there typically is little conscious recollection or even awareness of the details of procedural learning. Second, procedural learning is slow and incremental, and third it requires immediate and consistent feedback (Willingham, 1998).

Much evidence suggests that procedural learning is mediated largely within the striatum (Willingham, 1998; Mishkin et al., 1984; Saint-Cyr et al., 1988). Neuroanatomical studies suggest that the most important striatal regions might be the body and tail of the caudate nucleus and the putamen because these regions of the striatum receive direct projections from sensory association areas of cortex (Ashby et al., 2007).

As procedural learning requires many repetitions, it is not likely to influence performance when the categories have a simple structure that can be discovered quickly via hypothesis testing. But categories that cannot be learned via hypothesis testing are common in everyday life. For example, deciding whether an animal is a wolf or a German shepherd requires integrating information from a variety of perceptual dimensions in a way that is difficult to describe verbally. The laboratory analogue of this experience is the information-integration category-learning task, in which accuracy is maximized only if information from two or more stimulus dimensions is integrated at some pre-decisional stage. Typically, the optimal strategy is difficult or impossible to describe verbally (Ashby et al., 1998). An example is shown in Figure 5.1. In this case, each stimulus is a circular sine-wave grating that varies across trials in the width and orientation of the dark and light bars. The category boundary is denoted by the diagonal line. In this case, because of the incommensurable nature of the two stimulus dimensions, this bound is difficult (or impossible) to describe verbally. Even so, healthy young adults can reliably learn such categories (for a review, see Ashby & Maddox, 2005).

COVIS assumes that information-integration categories are primarily learned via striatal-mediated procedural learning (Ashby et al., 1998; Ashby & Waldron, 1999). In fact, there is strong evidence that information-integration tasks frequently recruit the
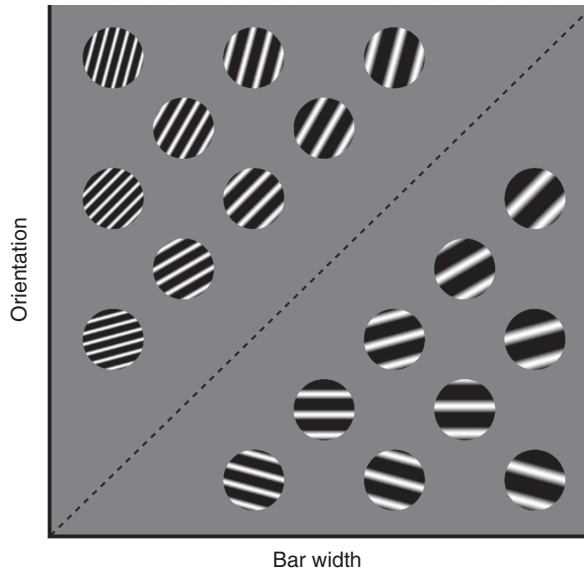
Fig. 5.1 Example of stimuli that might be used in an information-integration category-learning task. The diagonal line denotes the category boundary. Note that no simple verbal rule separates the discs into the two categories.

procedural learning system. The most popular paradigm for studying procedural learning is the serial-reaction-time (SRT) task (Nissen & Bullemer, 1987), in which subjects press keys as quickly as possible in response to stimuli that appear in various locations on the screen. A large response-time improvement occurs when the stimulus sequence is repeated, even when subjects are unaware that a sequence exists. In addition, changing the location of the response keys interferes with SRT learning, but changing the fingers that push the keys does not (Willingham et al., 2000). Thus, if procedural learning is used in information-integration tasks, then switching the locations of the response keys should interfere with learning, but switching the fingers that depress the keys should not. In fact, Ashby et al. (2003) reported evidence that directly supported this prediction. They also reported that neither manipulation had any effect on rule-based category learning. Subsequent studies have confirmed that information-integration learning requires a consistent association between a category and a response feature (e.g. the spatial location of the response) whereas rule-based learning does not require that this mapping is consistent (Maddox et al., 2004; Spiering & Ashby, 2008).

There is also evidence that as in traditional procedural-learning tasks, information-integration category learning is most effective when the feedback signal is delivered immediately after the categorization response. For example, delaying the feedback by as little as 2.5 s after the response significantly interferes with information-integration category learning, but delays as long as 10 s have no effect on rule-based learning (Maddox et al., 2003; see also Ashby et al., 1999,2002).

If information-integration tasks recruit striatal-dependent procedural learning, then information-integration learning should also depend on the striatum. In fact, there is good evidence that the striatum plays an important role in information-integration category learning. First, several studies have reported that Huntington's disease and Parkinson's disease patients are both impaired in difficult information-integration tasks (Filoteo et al., 2001a, 2004; Maddox & Filoteo, 2001), although other studies have reported that Parkinson's disease patients learn normally if the information-integration task is simpler (Ashby et al., 2003; Filoteo et al., 2004). Second, both known neuroimaging studies of information-integration category learning have reported significant task-related activation in the striatum (Nomura et al., 2007; Seger & Cincotta, 2002).

The perceptual representation memory system

The perceptual representation memory system (PRS) is a nondeclarative memory system that facilitates perceptual processing of a stimulus as a consequence of having seen that stimulus before (Schacter, 1990). Behavioural effects of the PRS can be observed after only a single stimulus repetition, yet are long lasting (e.g. Wiggs & Martin, 1998). Furthermore, PRS effects can be induced when two stimuli are different, but perceptually similar (e.g. Biederman & Cooper, 1992; Cooper et al., 1992).

The neural basis of the PRS is still unclear. The repetition priming thought to be mediated by the PRS is widely associated with the phenomenon of repetition suppression, in which repeated presentations of a stimulus elicit a smaller and smaller neural response in fMRI experiments. As a result, there have been specific proposals that repetition suppression is the neural signature of PRS activation (e.g. Schacter & Buckner, 1998; Wiggs & Martin, 1998). Despite this proposal, several open questions have prevented the development of a neural theory of the PRS. First, if the PRS is a purely perceptual memory system then we might expect to see its effects limited to sensory areas of cortex. Repetition suppression is not only often seen in the visual cortex, but it has also been reported in other nonsensory brain areas, including the PFC (e.g. Demb et al., 1995; Raichle et al., 1994). Second, the neural mechanisms that mediate repetition suppression are also unclear. For example, it is unclear whether repetition suppression is due to a sharpening of tuning curves (Wiggs & Martin, 1998) or because of rapid response learning (Dobbins et al., 2004).

This brief summary suggests that the PRS should be operating in any categorization experiment in which specific exemplars are shown repeatedly, or in which a category contains multiple exemplars that have high perceptual similarity. Of course, just because the PRS is activated does not mean that it aids the categorization process. To aid categorization, the PRS must give a different response when exemplars from contrasting categories are displayed. In its original description, Schacter (1990) argued that the PRS did not 'represent elaborative information that links an event to pre-existing knowledge' (p. 553). Instead, he proposed that the PRS could provide 'a basis for a feeling of familiarity' (p. 553). Thus, the PRS could assist in categorization only if exemplars from contrasting categories are associated with different levels of familiarity.

In most categorization experiments, PRS activation should be roughly equal for all categories. There is a popular exception though. In *prototype-distortion category-learning*

tasks, a category is created by first defining a category prototype and then by randomly distorting the prototype to create the other category members. In the most popular version, the prototype is a constellation of dots and the category exemplars are created by randomly perturbing the location of each dot in the prototype pattern (Posner & Keele, 1968). In the (*A*, not *A*) prototype-distortion task, there is a single prototype that is used to create category *A*. Stimuli not belonging to the *A* category are random patterns. The participant's task is to respond 'Yes' or 'No' depending on whether the presented stimulus was or was not a member of category *A*. In this task, the category *A* members have a coherent structure since they are created from a single prototype, but typically every pair of 'not *A*' category members are visually distinct. Thus, one might expect more PRS activation on *A* trials than on not *A* trials. In practice, this effect is often accentuated because it is common to include an initial training period when only category *A* members are displayed.

In contrast, in (*A*, *B*) prototype-distortion tasks, two prototypes are created and the *A* and *B* categories are constructed by randomly distorting these two patterns, respectively. An important feature of (*A*, *B*) tasks is that the stimuli associated with both responses each have a coherent structure. Thus, within-category similarity is equally high in both categories, so it is natural to predict approximately equal amounts of PRS activation on *A* and *B* trials. Historically, prototype-distortion tasks have been run both in (*A*, *B*) and (*A*, not *A*) forms, although (*A*, not *A*) tasks are most common.

There have been several proposals that the PRS mediates much of the learning that occurs in the (*A*, not *A*) prototype-distortion task (Ashby & O'Brien, 2005; Casale & Ashby, 2008; Reber & Squire, 1999), and behavioural support for this hypothesis was provided by Casale and Ashby (2008). Neuropsychological support comes from studies showing that a variety of patient groups with known deficits in rule-based and information-integration category learning are apparently normal in (*A*, not *A*) prototype-distortion learning. This includes patients with Parkinson's disease (Reber & Squire, 1999), schizophrenia (Kéri et al., 2001), or Alzheimer's disease (Sinha, 1999). Normal (*A*, not *A*) performance has also been shown in patients with amnesia (Knowlton & Squire, 1993; Kolodny, 1994; Squire & Knowlton, 1995). At least two studies have compared (*A*, not *A*) and (*A*, *B*) prototype-distortion learning on the same patients – and both studies report the same striking dissociation. Specifically, Sinha (1999) reported normal (*A*, not *A*) performance in Alzheimer's disease patients, but impaired (*A*, *B*) performance, and Zaki et al. (2003) reported this same pattern of results with amnesiacs. Sinha (1999) also reported deficits in (*A*, *B*) prototype-distortion learning in patients with amnesia.

A handful of neuroimaging studies have used prototype-distortion tasks. All of these using an (*A*, not *A*) task have reported learning-related changes in occipital cortex (Aizenstein et al., 2000; Reber et al., 1998) – in general, reduced occipital activation was found in response to category *A* exemplars, although Aizenstein et al. (2000) found this reduction only under implicit learning conditions. When subjects were given explicit instructions to learn the *A* category, increased occipital activation was observed. Studies that used (*A*, *B*) tasks have reported quite different results. Seger et al. (2000) did report categorization-related activation in occipital cortex, but they also found significant learning-related changes in prefrontal and parietal cortices. Vogels et al. (2002)

reported results from a hybrid task in which subjects were to respond *A*, *B*, or *neither*. Thus, stimuli were created either from distortions of an *A* prototype, a *B* prototype, or were just random patterns. Like Seger et al. (2000), Vogels et al. (2002) found prefrontal and parietal activation (although in different foci). However, they also reported task-related activation in orbitofrontal cortex and the striatum, and they failed to find any task-related activation in occipital cortex.

Automatic categorization

The previous section focused on category learning. However, there are many reasons to believe that the neural mechanisms and pathways that mediate category learning are different from the neural structures that mediate automatic responses to highly learned categories. For example, many neuropsychological groups that are impaired in category learning (e.g. frontal patients and Parkinson's disease patients) do not lose old, familiar categories (e.g. fruits and tools). Similarly, there is no evidence that people who lose a familiar category (i.e. who develop a category-specific agnosia) develop any general category-learning deficits (although we know of no studies that directly address this issue). As another example, there is strong evidence, much of which we reviewed in the previous section, that certain types of category learning (e.g. information-integration) are heavily dependent on dopamine. Some recent results however, indicate that instrumental behaviours that are dopamine dependent during early learning become dopamine independent after extensive training (i.e. weeks).

Neuroscience data

Compared to the amount of research that has been done on category learning, relatively few studies have examined the neural basis of automatic categorization responses. In one of the few single-unit recording studies to examine this issue, Muhammad et al. (2006) recorded from single neurons in the PFC, head of the caudate, and premotor cortex while monkeys were making rule-based categorization responses. As predicted by COVIS, they found many cells in the PFC and caudate nucleus that fired selectively to a particular rule. However, after training the animals for a year, they also found many premotor cells that were rule selective, and even more importantly, these cells responded on average about 100 ms *before* the PFC rule-selective cells. Thus, after categorization had become automatic, the PFC, although still active, was not mediating response selection.

A similar result was reported by Carelli et al. (1997). In this study, rats were trained to lever press to a tone. Thus, Carelli et al. (1997) studied an instrumental behaviour, but not categorization. The animals learned to lever press to the tone within just a few sessions, but they received many extra days of overtraining. Throughout this extended training period, Carelli et al. (1997) recorded from single units in the striatum. As expected, just after learning, many units were found that fired a burst just before the animal lever pressed. However, a few sessions later, the same striatal units still fired bursts, but now these bursts came *after* the response had been made, and therefore they played no role in response selection. In later sessions, presumably after automaticity was well established, the striatum quit responding altogether – that is, neither

the tone nor the response elicited any activity from the same striatal units that apparently controlled the response earlier in training.

Choi et al. (2005) reported the results of an experiment that used a task similar to the lever-pressing task of Carelli et al. (1997). However, rather than recording from single units, each animal was injected either with a saline control or with a selective dopamine D1 antagonist at some point during training. As expected, immediately after initial learning was complete, the D1 antagonist significantly interfered with the expression of the learned behaviour in a dose-dependent manner. In contrast, after a period of significant overtraining, the D1 antagonist had no effect on the performance of the learned behaviour.

Other results indicate that during initial acquisition of a more traditional procedural learning task (i.e. the SRT task), task-related activation in the striatum is mostly in the caudate nucleus, whereas after extended training, the task-related activation shifts to the putamen (e.g. Lehéricy et al., 2005; Miyachi et al., 2002). As with the Carelli et al. (1997) and Choi et al. (2005) data, these results are problematic for any model that assumes instrumental learning is controlled by a single basal ganglia-dependent system.

We know of no neuroimaging studies that have specifically investigated automatic categorization. However, several neuroimaging studies have examined changes in neural activations during the course of more traditional procedural-learning tasks in which participants were required to execute a fixed sequence of finger movements. In general, automaticity was associated with reduced neural activation. For example, after automaticity was achieved, decreased activation (relative to initial learning) was reported in cingulate, premotor, parietal, and prefrontal cortices, as well as in the caudate nucleus (Lehéricy et al., 2005; Poldrack et al., 2005; Wu et al., 2004).

There are many reports of neuropsychological patients who seem to lose a highly specific and very familiar category (Warrington & Shallice, 1984), a condition known as category-specific agnosia. For example, a variety of patients have been identified who have near-normal categorization performance with inanimate objects, but are highly impaired in their ability to categorize living things (e.g. Satori & Job, 1988). The most widely known of such conditions is prosopagnosia, which is a category-specific agnosia that is limited to faces. Category-specific agnosias, including prosopagnosia, often occur following a lesion to visual cortex. Single-unit recording and neuroimaging studies have identified inferotemporal cortex (and especially the fusiform gyrus) as a region especially critical to the object-perception process (e.g. McCarthy et al., 1997).

A separate literature shows how experience with a category affects the representation of its members in visual cortex (e.g. Humphreys & Forde, 2001). For example, in several studies categorization training enhanced the sensitivity of cells in the inferotemporal cortex to diagnostic features compared to features that were irrelevant to the categorization judgement (e.g. Sigala & Logothetis, 2002). Note that such changes are consistent with the widely held view that category learning is often associated with changes in the allocation of perceptual attention (Nosofsky, 1986).

Despite these many results indicating a significant role for visual cortex in the categorization process, there is good reason to believe that category learning is not

mediated within visual cortex. For example, when the motor responses associated with category membership are switched (i.e. from 'approach' to 'avoid' and vice versa), the firing properties of cells in the inferotemporal cortex that are sensitive to those stimuli do not change (Rolls et al., 1977). More recent studies have found similar null results in traditional categorization tasks (e.g. Freedman et al., 2003; Op de Beeck et al., 2001). In each case, single-cell recordings showed that the firing properties of cells in the inferotemporal cortex did not change as monkeys learned to classify visual objects into one of two categories. The cells showed sensitivity to specific visual images, but category training did not make them more likely to respond to other stimuli in the same category, or less likely to respond to stimuli belonging to the contrasting category. For these reasons, the best evidence seems to suggest that inferotemporal cortex does not mediate the learning of new categories. Even so, this visual-association area is crucial to the categorization process, because it appears to encode a high-level representation of the visual stimulus.

Theories of automatic categorization

In response to these differences between category learning and automatic categorization, Ashby et al. (2007) proposed a biologically detailed computational model of how categorization judgements become automatic in tasks that depend on procedural learning. The model assumes that there are two neural pathways from the relevant sensory-association area to the premotor area that mediates response selection. The longer and slower path, which is identical to the procedural learning system of COVIS, projects from the sensory-association cortex to the premotor cortex via the striatum, globus pallidus, and thalamus. A faster, purely cortical path projects directly from the sensory-association area to the premotor area. The model assumes that the subcortical path, although slower, has greater neural plasticity because it receives a dopamine-mediated learning signal from the substantia nigra. In contrast, the faster cortical–cortical path learns more slowly via (dopamine-independent) classical two-factor Hebbian learning. Due to its greater plasticity, early learning is dominated by the subcortical path, but the development of automaticity is characterized by a transfer of control to the faster cortical–cortical projection.

Although we know of no other neural theories of categorization automaticity, there are proposals that as training progresses, control of procedural-learning tasks passes from the caudate to the putamen (e.g. Costa, 2007). A theory that the development of categorization automaticity is mediated by such a shift accounts for much of the data reviewed in this section. Even so, it is not clear how such a theory could account for the normal automatic categorization behaviour of Parkinson's patients, since the effects of Parkinson's disease are particularly insidious in the putamen. In addition, because the putamen is a prominent target of the nigrostriatal dopamine system, it is not clear how this theory could account for results (e.g. Choi et al., 2005) showing that overlearned behaviours are largely dopamine independent.

Conclusions

Ten years ago, virtually nothing was known about the neurobiology of category learning and the dominant view was that humans had a single category-learning system.

During the past decade, overwhelming evidence has been collected supporting the multiple-systems viewpoint and much has been learned about the brain regions and neural pathways that mediate learning in these putative systems. During the coming years, it is likely that this trend will continue. We can expect more details to emerge about the neural mechanisms that subserve each system, and about how the various systems interact during the learning process.

One surprising result of this new cognitive-neuroscience focus on category learning is that virtually every such study has implicated the striatum. We believe there are several reasons for this. One is that the striatum plays a key role in both working memory (i.e. the head of the caudate nucleus) and procedural memory (i.e. body and tail of caudate and putamen), and the evidence reviewed above suggests that one of these two memory systems is active in virtually every category-learning situation. A second related reason is that the striatum may be the only brain region capable of reinforcement learning (Ashby et al., 2007) – that is, learning in which synapses active on trials when correct feedback is given are strengthened and synapses active on trials when error feedback is given are weakened. Whereas it might not be surprising that reinforcement learning is critical in information-integration tasks, the striatal activation observed by Seger and Cincotta (2005) in a task with unstructured categories suggests that reinforcement learning, and hence the striatum, may play an important role in a wide variety of category-learning paradigms.

The cognitive neuroscience efforts also have led directly to the discovery of a variety of important new empirical phenomena. Here we mention only one (for more examples, see Ashby & Maddox, 2005). Neuroscience evidence suggests that the trace (i.e. Ca^{2+}) of an active synapse persists for only a few seconds after a striatal medium spiny neuron fires. For this reason, COVIS (Ashby et al., 1998) predicts that information-integration learning should therefore be impaired if the feedback signal is delayed by several seconds, and also that such delays should not affect rule-based learning. These predictions were both supported by Maddox et al. (2003). It is difficult to imagine a scenario in which a researcher working within the cognitive tradition would ever think of running this experiment. Certainly, none of the current or past cognitive models of categorization offer any hint that feedback delay is an interesting independent variable to study.

Cognitive neuroscience research on categorization also offers the hope of unifying the historically separate human and animal literatures. Although the cognitive-neuroscience work has focused on human adults, all mammals have a cortex and basal ganglia, and thus, they are likely to have the same memory systems as humans. But the greatest differences between the neuroanatomy of humans and other mammals are in frontal cortex, and more specifically in the size and complexity of the PFC. Thus, although the system that humans rely on in rule-based tasks may exist in some rudimentary form in all mammals, one might expect that this system is considerably more primitive in species further down the phylogenetic scale. Healthy, young human adults learn rule-based tasks approximately ten times faster than they do an information-integration task with equally coherent and separated categories. We predict this difference is greatest in humans, and should gradually reduce across species as the size and complexity of the PFC decreases.

Finally, one of the unexpected benefits of this new focus on the cognitive neuroscience of categorization is that it has tied the categorization process more closely to other cognitive tasks. For example, because the striatum was implicated in information-integration category learning, previously unknown relationships were discovered between information-integration categorization and more traditional procedural-learning tasks (e.g. the SRT task).

Acknowledgements

References

Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., et al. (2000). Complementary category learning systems identified using event-related functional MRI. *Journal of Cognitive Neuroscience*, **12**, 977–987.

Amos, A. (2000). A computational model of information processing in the frontal cortex and basal ganglia. *Journal of Cognitive Neuroscience*, **12**, 505–519.

Asaad, W. F., Rainer, G., & Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology*, **84**, 451–459.

Ashby, F. G. & Ennis, J. M. (2006). The role of the basal ganglia in category learning. *The Psychology of Learning and Motivation*, **46**, 1–36.

Ashby, F. G. & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, **56**, 149–178.

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

Ashby, F. G. & Valentin, V. V. (2005). Multiple systems of perceptual category learning: Theory and cognitive tests. In H. Cohen & C. Lefebvre (Eds.), *Categorization in Cognitive Science*, pp. 543–578, New York: Elsevier.

Ashby, F. G. & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin and 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 and Psychophysics*, **61**, 1178–1199.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory and Cognition*, **30**, 665–676.

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

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

Ashby, F. G., Ell, S. W., Valentin, V. V., & Casale, M. B. (2005). FROST: A distributed neurocomputational model of working memory maintenance. *Journal of Cognitive Neuroscience*, **17**, 1728–1743.

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization, *Psychological Review*, **114**, 632–656.

Baddeley, A. D. (1986). *Working Memory* New York: Oxford University Press.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition*, **11**, 211–227.

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

Biederman, I. & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, **18**, 121–133.

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

Carelli, R. M., Wolske, M., & West, M. O. (1997). Loss of lever press-related firing of rat striatal forelimb neurons after repeated sessions in a lever pressing task. *Journal of Neuroscience*, **17**, 1804–1814.

Casale, M. B. & Ashby, F. G. (2008). A role for the perceptual representation memory system in category learning. *Perception and Psychophysics*, **70**, 983–999.

Choi, W. Y., Balsam, P. D., & Horvitz, J. C. (2005). Extended habit training reduces dopamine mediation of appetitive response expression. *Journal of Neuroscience*, **25**, 6729–6733.

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, L. A., Schacter, D. L., Ballesteros, S., & Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, 43–57.

Costa, R. M. (2007). Plastic corticostriatal circuits for action learning. *Annals of the New York Academy of Science*, **1104**, 172–191.

Crone, E. A., Wendelken, C., Donohue, S. E., & Bunge, S. A. (2005). Neural evidence for dissociable components of task-switching. *Cerebral Cortex*, **16**, 475–486.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, **107**, 284–294.

Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. E. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *Journal of Neuroscience*, **15**, 5870–5878.

Dobbins, I. G., Schnyer, D. M., Verfaellie, M., & Schacter, D. L. (2004). Cortical activity reductions during repetition priming can result from rapid response learning. *Nature*, **428**, 316–319.

Eichenbaum, H. (1997). Declarative memory: Insights from cognitive neurobiology. *Annual Review of Psychology*, **48**, 547–572.

Elliott, R., Rees, G., & Dolan, R. J. (1999). Ventromedial prefrontal cortex mediates guessing. *Neuropsychologia*, **37**, 403–411.

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

Filoteo, J. V., Maddox, W. T., and Davis, J. D. (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.

Filoteo, V. J., Maddox, W. T., Salmon D. P., & Song D. D. (2007). Implicit category learning performance predicts rate of cognitive decline in nondemented patients with Parkinson's disease. *Neuropsychology*, **21**, 183–192.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience*, **1**, 137–160.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, **23**, 5235–5246.

Fuster, J. M. (1989). *The Prefrontal Cortex*, 2nd edn., Philadelphia, PA: Lippincott-Raven.

Heaton, R. K. (1981). *A Manual for the Wisconsin Card Sorting Test* Odessa, FL: Psychological Assessment Resources.

Heimer, L. (1995). *The Human Brain and Spinal Cord*, 2nd edn., New York: Springer-Verlag.

Hikosaka, O., Sakamoto, M., & Sadanari, U. (1989). Functional properties of monkey caudate neurons III. Activities related to expectation of target and reward. *Journal of Neurophysiology*, **61**, 814–831.

Hopkins, R. O., Myers, E. C., Shohamy, D., Grossman, S., & Gluck, M. (2004). Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia*, **42**, 524–535.

Hoshi, E., Shima, K., & Tanji, J. (1998). Task-dependent selectivity of movement-related neuronal activity in the primate prefrontal cortex. *Journal of Neurophysiology*, **80**, 3392–3397.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia*, pp. 249–270, Cambridge, MA: MIT Press.

Humphreys, G. W. & Forde, E. M. E. (2001). Hierarchies, similarity and interactivity in object recognition: 'Category-specific' neuropsychological deficits. *Behavioral and Brain Sciences*, **24**, 453–509.

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.

Jaspers, R. M., De Vries, T. J., & Cools, A. R. (1990). Enhancement in switching motor patterns following local application of the glutamate agonist AMPA into the cat caudate nucleus. *Behavioral Brain Research*, **37**, 237–246.

Joel, D., Weiner, I., & Feldon, J. (1997). Electrolytic lesions of the medial prefrontal cortex in rats disrupt performance on an analog of the Wisconsin Card Sorting Test, but do not disrupt latent inhibition: Implications for animal models of schizophrenia. *Behavioral Brain Research*, **85**, 187–201.

Kéri, S., Kelemen, O., Benedek, G., & Janka, Z. (2001). Intact prototype learning in schizophrenia. *Schizophrenia Research*, **52**, 261–264.

Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1997). Effects of bromocriptine on human subjects depends on working memory capacity. *Neuroreport*, **8**, 3581–3585.

Kincaid, A. E., Zheng, T., & Wilson, C. J. (1998). Connectivity and convergence of single corticostriatal axons. *Journal of Neuroscience*, **18**, 4722–4731.

Knowlton, B. J. & Squire, L. R. (1993). The learning of natural categories: Parallel memory systems for item memory and category-level 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.

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.

Konishi, S., Karwazu, M., Uchida, I., Kikyo, H., Asakura, I., & Miyashita, Y. (1999). Contribution of working memory to transient activation in human inferior prefrontal cortex during performance of the Wisconsin Card Sorting Test. *Cerebral Cortex*, **9**, 745–753.

Lea, S. E. G. (1984). In what sense do pigeons learn concepts? In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal Cognition*, pp. 263–277, Hillsdale, NJ: Erlbaum.

Lehéricy. S., Benali, H., Van de Moortele, P. F., et al. (2005). Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proceedings of the National Academy of Science*, **102,** 12566–12571.

Leng, N. R. & Parkin, A. J. (1988). Double dissociation of frontal dysfunction in organic amnesia. *British Journal of Clinical Psychology*, **27**, 359–362.

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.

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. (2004). Evidence for a procedural learning-based system in perceptual category learning. *Psychonomic Bulletin and Review*, **11**, 945–952.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, **9**, 605–610.

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

Merchant, H., Zainos, A. Hernadez, A., Salinas, E., & Romo, R. (1997). Functional properties of primate putamen neurons during the categorization of tactile stimuli. *Journal of Neurophysiology*, **77**, 1132–1154.

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.

Miyachi, S., Hikosaka, O. & Lu, X. (2002). Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Experimental Brain Research*, **146**, 122–126.

Monchi, O., Taylor, J. G., & Dagher, A. (2000). A neural model of working memory processes in normal subjects, Parkinson's disease, and schizophrenia for fMRI design and predictions. *Neural Networks*, **13**, 953–973.

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.

Muhammad, R., Wallis, J. D., & Miller, E. K. (2006). A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. *Journal of Cognitive Neuroscience*, **18**, 1–16.

Nissen, M. J. & Bullemer, P. (1987). Attention requirements of learning: Evidence from performance measures. *Cognitive Psychology*, **19**, 1–32.

Nomura, E. M., Maddox, W. T., Filoteo, J. V., et al. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, **17***, 37–43.

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

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory and Cognition*, **22**, 352–369.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53–79.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of paramterized shapes. *Nature Neuroscience*, **4**, 1244–1252.

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. *Behavior Neuroscience*, **106**, 439–446.

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.

Poldrack, R. A., Sabb, F., Foerde, K., et al. (2005). The neural correlates of automaticity during motor skill learning. *Journal of Neuroscience*, **25**, 5356–5364.

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 Neuroscience*, **13**, 25–42.

Raichle, M. E., Fiez, J. A., Videen, T. O., et al. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex*, **4**, 8–26.

Reber, P. J. & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. *Behavioral Neuroscience*, **113**, 235–242.

Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 747–750.

Roberts, A. C., De Salvia, M. A., Wilkinson, L. S., et al. (1994). 6-hydroxydopamine lesions of the prefrontal cortex in monkeys enhance performance on an analog of the Wisconsin card sort test: Possible interactions with subcortical dopamine. *Journal of Neuroscience*, **14**, 2531–2544.

Roberts, A. C. & Wallis, J. D. (2000). Inhibitory control and affective processing in the prefrontal cortex: Neuropsychological studies in the common marmoset. *Cerebral Cortex*, **10**, 252–262.

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.

Rolls, E. T., Judge, S. J., & Sanghera, M. (1977). Activity of neurons in the inferotemporal cortex of the alert monkey. *Brain Research*, **130**, 229–238.

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

Sakamoto, Y. & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, **133**, 534–553.

Satori, G. & Job, R. (1988). The oyster with four legs: A neuropsychological study on the interaction of visual and semantic information. *Cognitive Neuropsychology*, **5**, 105–132.

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

Schacter, D. L. & Buckner, R. L. (1998). Priming and the brain. *Neuron,* **20**, 185–195.

Schultz, W. & Romo, R. (1992). Role of primate basal ganglia and frontal cortex in the internal generation of movements. I. Preparatory activity in the anterior striatum. *Experimental Brain Research*, **91**, 363–384.

Schultz, W., Dayan, P. & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, **275**, 1593–1599.

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

Seger, C. A. & Cincotta, C. M. (2002). Striatal activity in concept learning. *Cognitive, Affective, and Behavioral Neuroscience*, **2**, 149–161.

Seger, C. A. & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience*, **25**, 2941–2951.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, **75**, (13, Whole No. 517).

Sigala, N. & Logothetis, N. K. (2002). Visual categorization shape feature selectivity in the primate temporal cortex. *Nature*, **415**, 318–320.

Sinha, R. R. (1999). Neuropsychological substrates of category learning. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, **60**(5-B), 2381. (UMI No. AEH9932480).

Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in Rhesus monkeys: A study of the Shepard, Hovland, and Jenkins tasks. *Journal of Experimental Psychology*: *General*, **133**, 398–414.

Spiering, B. J. & Ashby, F. G. (2008). Response processes in information-integration category learning. *Neurobiology of Learning and Memory,* **90**, 330–338.

Squire, L. R. & Knowlton, B. J. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Science USA*, **92,** 12470–12474.

Squire, L. R. & Schacter, D. L. (2002). *The Neuropsychology of Memory*, 3rd edn., New York: Guilford Press.

Sutton, R. S. & Barto, A. G. (1998). *Reinforcement Learning* Cambridge, MA: MIT Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, **53**, 1–25.

van Domburg, P. H. & ten Donkelaar, H. J. (1991). *The Human Substantia Nigra and Ventral Tegmental Area* Berlin: Springer-Verlag.

Van Golf Racht-Delatour, B. & El Massioui, N. (1999). Rule-based learning impairment in rats with lesions to the dorsal striatum. *Neurobiology of Learning and Memory*, **72**, 47–61.

Vogels, R., Sary, G., Dupont, P., & Orban, G. A. (2002). Human brain regions involved in visual categorization. *Neuroimage*, **16**, 401–414.

Waldron, E. M. & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin and Review*, **8**, 168–176.

Warrington, E. K. & Shallice, T. (1984). Category specific semantic impairments. *Brain*, **107**, 829–854.

White, I. M. & Wise, S. P. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Experimental Brain Research*, **126**, 315–335.

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

Wiggs, C. L. & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, **8**, 227–233.

Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological 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. 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, MA: Bradford.

Wu, T., Kansaku, K. & Hallett, M. (2004). How self-initiated memorized movements become automatic: A functional MRI study. *Journal of Neurophysiology*, **91**, 1690–1698.

Zaki, S. R., Nosofsky, R. M., Jessup, N. M., & Unversagt, F. W. (2003). Categorization and recognition performance of a memory-impaired group: Evidence for single-system models. *Journal of the International Neuropsychological Society*, **9**, 394–406.

Zeithamova, D. & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory and Cognition*, **34**, 387–398.