

Multiple Systems of Perceptual Category Learning: Theory and Cognitive Tests

F. Gregory Ashby and Vivian V. Valentin

University of California, Santa Barbara, CA, United States

OUTLINE

7.1 Introduction	158
7.2 Rule-Based Versus Information-Integration Category Learning Tasks	159
7.3 COVIS	162
7.3.1 The COVIS Declarative System	162
7.3.2 The COVIS Procedural-Learning System	167
7.3.3 Dissociations Between RB and II Categorization	174
7.4 Conclusions	178
Acknowledgments	180
References	180

Abstract

The first neurobiologically detailed theory of multiple systems in category learning, called COVIS, was originally conceived in 1998. COVIS, which is now well established, postulates two systems that compete throughout learning—a frontal-based declarative system that uses logical reasoning and depends on working memory and executive attention, and a basal ganglia-mediated system that uses procedural-learning. The procedural system can learn a wide variety of category

structures, but it learns in a slow incremental fashion and is highly dependent on reliable and immediate feedback. In contrast, the declarative rule-based (RB) system can learn a fairly small set of category structures quickly—specifically, those structures that can be learned via a logical reasoning process. These two systems learn simultaneously, but as long as RB strategies lead to successful performance, the declarative system inhibits the procedural system. This theory is described in detail and a variety of cognitive behavioral and cognitive neuroscience experiments are reviewed that test some parameter-free a priori predictions made by COVIS.

7.1 INTRODUCTION

All animals must learn to categorize objects and events in their environment to survive. Is the mushroom edible or poisonous? Is the object behind the bush a deer or a wolf? Correct classification allows animals to select the appropriate approach or avoidance response to nutrients and poisons, and to prey and predators. Given the importance of categorization, it is not surprising that many different cognitive theories of human category learning have been proposed and tested. Although some of these theories have unquestionably been more successful than others, most categorization researchers would probably acknowledge disappointment in how difficult it has been to differentiate the predictions of the more successful models, despite the very different cognitive processes they hypothesize. It appears that behavioral category learning data do not offer enough constraints to identify the correct model.

The cognitive neuroscience revolution has offered some exciting new tools for resolving these conflicts. In particular, the past few decades has seen an explosion of new results that collectively are beginning to paint a detailed picture of the neural mechanisms and pathways that mediate category learning. These results come from a wide variety of sources, including human behavioral, animal lesion, single-cell recording, neuroimaging, and neuropsychological patient studies. Lagging somewhat behind this avalanche of new data has been the development of new theories that can account for the traditional cognitive results as well as for these newer neuroscience results. Even so, some such theories have been proposed. The most comprehensive and best tested of these is the COVIS theory of category learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Crossley, 2011; Ashby, Ennis, & Spiering, 2007; Ashby & Waldron, 1999; Cantwell, Crossley, & Ashby, 2015). This chapter describes COVIS and many of the cognitive experiments that have been run to test some strong and surprising a priori

predictions of COVIS. The next chapter, entitled *The Neuropsychology of Perceptual Category Learning*, describes tests of COVIS using neuropsychological patient data.

Briefly, COVIS postulates two systems that compete throughout learning—a frontal-based system that learns explicit rules and depends on declarative memory systems and a basal ganglia-mediated procedural-learning system. The procedural system is phylogenetically older. It can learn a wide variety of category structures, but it learns in a slow incremental fashion and is highly dependent on reliable and immediate feedback. In contrast, the declarative rule learning system can learn a fairly small set of category structures quickly—specifically, those structures in which the contrasting categories can be separated by simple explicit rules. Tasks that require subjects to learn such structures are called rule-based (RB) category learning tasks. On the other hand, there are many category structures that the declarative system cannot learn. An important example occurs in information-integration (II) tasks, in which learning requires subjects to integrate perceptual information across two or more incommensurable stimulus dimensions. Before describing COVIS in detail, we take a short detour to describe RB and II category learning tasks.

7.2 RULE-BASED VERSUS INFORMATION-INTEGRATION CATEGORY LEARNING TASKS

There is now abundant evidence that declarative and procedural memory systems both contribute to perceptual category learning (e.g., [Ashby & Maddox, 2005, 2010](#); [Eichenbaum & Cohen, 2001](#); [Poldrack et al., 2001](#); [Poldrack & Packard, 2003](#); [Squire, 2004](#)). Much of this evidence comes from RB and II category learning tasks. In RB tasks, the categories can be learned via some logical reasoning process ([Ashby et al., 1998](#)). In the most common applications, only one stimulus dimension is relevant, and the participant's task is to discover this relevant dimension and then to map the different dimensional values to the relevant categories. Meanwhile the variations in the irrelevant dimension must be ignored. However, there is no requirement that RB tasks be one-dimensional. For example, a conjunction rule (e.g., respond A if the stimulus is small on dimension x and small on dimension y) is an RB task because a conjunction is a pair of logical conditionals. There can also be complex logical rules in which only one dimension is relevant, such as the disjunction (e.g., respond A if the stimulus is small or large, and B for intermediate values on dimension x). The key is that separate decisions are first made about each relevant dimension and then these separate decisions

are combined following the rules of Boolean algebra. A variety of evidence suggests that success in RB tasks depends on declarative memory systems and especially on working memory and executive attention (Ashby et al., 1998; Maddox, Ashby, Ing, & Pickering, 2004; Waldron & Ashby, 2001; Zeithamova & Maddox, 2006).

In II category learning tasks, accuracy is maximized only if information from two or more incommensurable stimulus components is integrated at some predecisional stage (Ashby & Gott, 1988; Ashby et al., 1998). Perceptual integration could take many forms—from computing a weighted linear combination of the dimensional values to treating the stimulus as a Gestalt. Typically, the optimal strategy in II tasks is difficult or impossible to describe verbally. Real-world examples of II tasks are common. For example, deciding whether an x-ray shows a tumor requires years of training and expert radiologists are only partially successful at describing their categorization strategies. Evidence suggests that success in II tasks depends on procedural-learning that is mediated largely within the striatum (Ashby & Ennis, 2006; Filoteo, Maddox, Salmon, & Song, 2005; Knowlton, Mangels, & Squire, 1996; Nomura et al., 2007).

Fig. 7.1 shows typical examples of RB and II tasks. In both cases, the two categories are composed of circular sine-wave gratings (i.e., discs in which luminance varies sinusoidally). The discs are all of the same size, shape, and contrast, but they vary in the narrowness and orientation of the dark and light bars. The solid lines denote the category boundaries. In most experiments, each category contains hundreds of these discs. On each trial of a typical experiment, a disc is randomly selected and presented to the subject, whose task is to assign it to category A or B. In most cases, feedback about response accuracy is given at the end of each trial, and this process is repeated for hundreds or thousands of trials. Note that a simple verbal rule perfectly partitions the categories in the RB task (“Respond A if the bars are narrow and B if the bars are wide”), but no verbal rule correctly separates the two categories in the II task. For a more thorough discussion of RB and II tasks, see Ashby and Maddox (2005).

The Fig. 7.1 RB and II categories are essentially just 45° rotations of each other. This means they are almost perfectly matched on all category separation statistics. As a result, any standard cluster analysis routine would perform identically on the two category structures. Even so, healthy human adults learn the RB categories much faster than the II categories (e.g., Ashby & Maddox, 2005). COVIS predicts this is because humans discover the optimal strategy in the RB task explicitly, whereas an explicit, hypothesis-testing approach fails in the II task, so instead, control must be passed to the phylogenetically older,

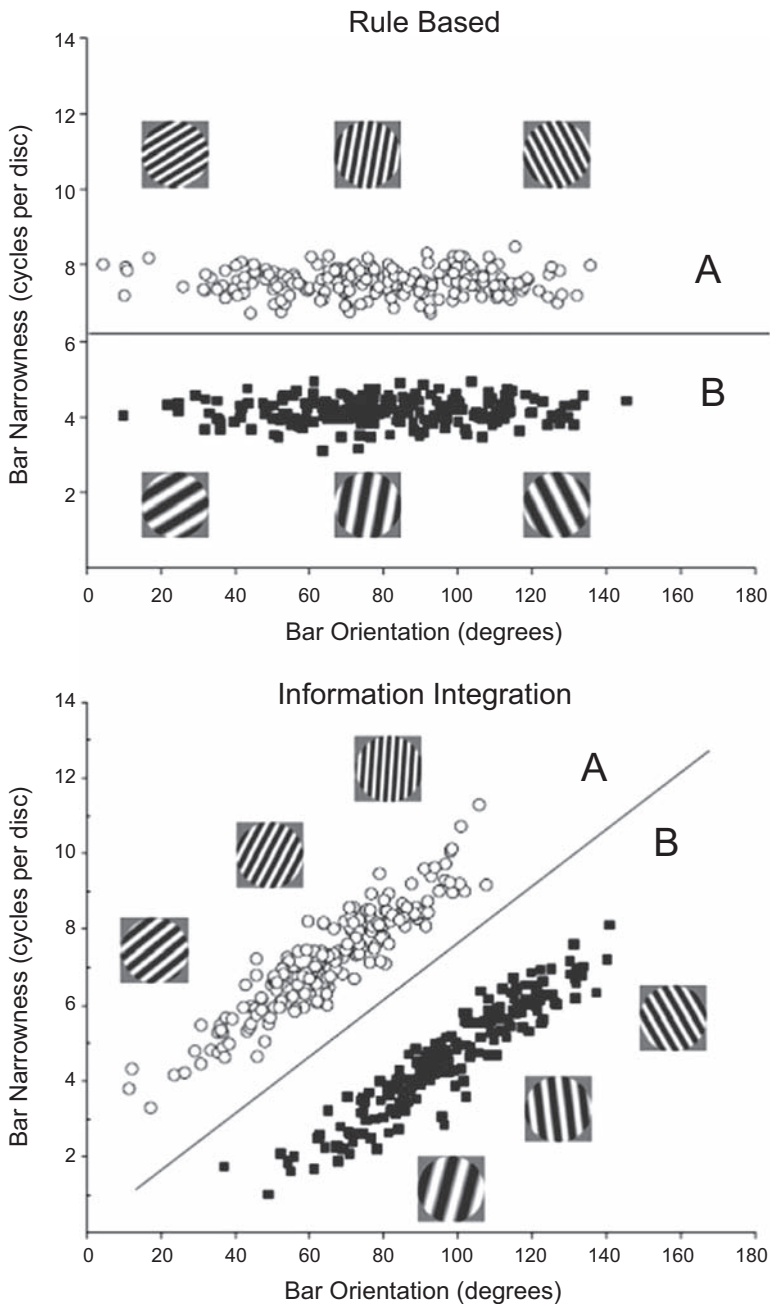


FIGURE 7.1 Examples of rule-based (RB) and information-integration (II) category structures. Each stimulus is a sine-wave disc that varies across trials in bar narrowness and bar orientation. For each task, three illustrative Category A and B stimuli are shown. The small rectangles and open circles denote the specific values of all stimuli used in each task. In the RB task, only bar narrowness carries diagnostic category information, so the optimal strategy is to respond with a one-dimensional bar-narrowness rule (thick versus thin), while ignoring the orientation dimension. In the II task, both bar narrowness and orientation carry useful but insufficient category information. The optimal strategy requires integrating information from both dimensions in a way that is impossible to describe verbally.

procedural-learning system. An alternative view, however, is that the II task is inherently more complex (Nosofsky, Stanton, & Zaki, 2005). A strong test between these two hypotheses would be to run the Fig. 7.1 RB and II tasks on a cognitively simpler species, such as pigeons. If the II task is inherently more complex then the performance difference should be even greater in pigeons than in humans, because tasks that humans struggle with should be even more of a challenge for a cognitively simpler species. In contrast, if the human RB advantage is because humans learn the RB task explicitly then the performance difference should be reduced in a species that lacks a well-developed explicit reasoning ability because such a species would have to learn the two category structures in the same way. In fact, pigeons learn both category structures equally well and at exactly the same rate (Smith et al., 2011), which is strong evidence that the II task is not inherently more complex than the RB task and therefore that the human RB advantage is because humans learn these two category structures in fundamentally different ways.

Like humans, both macaque and capuchin monkeys also learn the Fig. 7.1 RB categories more quickly than the II categories (Smith et al., 2012). This result suggests that the human RB learning advantage is not necessarily language based. Another (speculative) possibility is that the critical attribute supplied by declarative memory systems is executive, selective attention—a skill that is closely tied to prefrontal cortex (e.g., Miller & Cohen, 2001). Identifying the mechanisms that allow humans (and monkeys) to learn RB categories so quickly is an important topic of future research.

7.3 COVIS

As mentioned earlier, COVIS postulates two systems that compete throughout learning—a declarative memory system that learns explicit rules and a procedural memory system that uses a form of implicit and incremental learning. This section describes those systems in detail.

7.3.1 The COVIS Declarative System

The COVIS declarative system assumes subjects generate and test explicit hypotheses about category membership. For example, the initial hypothesis may be to “respond A if the grating is tilted up, and B if it is tilted down.” This candidate rule is then held in working memory while it is being tested. With the RB categories shown in Fig. 7.1,

performance based on this rule will be at chance, so the feedback will signal the subject that this hypothesis is incorrect. At this point, an alternative hypothesis must be selected, and executive attention must be switched from the old to the new hypothesis. These selection, switching, and testing procedures continue until performance is satisfactory, or until the subject gives up and decides that no satisfactory rule exists.

Fig. 7.2 shows the neural structures that mediate performance in the COVIS declarative system during a trial of the category learning tasks illustrated in Fig. 7.1. The key structures in the model are the anterior cingulate, the prefrontal cortex (PFC), the head of the caudate nucleus, and the hippocampus. There are two separate subnetworks in this model—one that maintains candidate rules in working memory, tests those rules, and mediates the switch from one rule to another, and another that generates or selects new candidate hypotheses.

The working memory maintenance and attentional switching network includes all structures in Fig. 7.2, except the anterior cingulate. The idea is that the long-term representation of each possible salient rule is encoded in some neural network in sensory association cortex. These cortical units send excitatory signals to working memory units in lateral PFC, which send recurrent excitatory signals back to the same

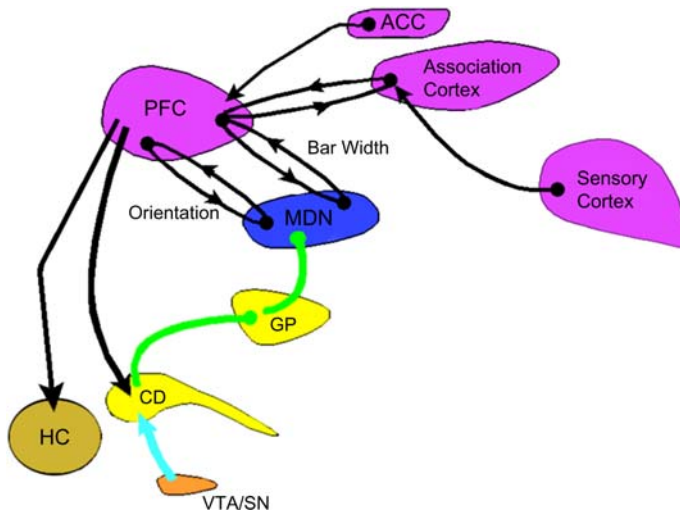


FIGURE 7.2 The COVIS declarative system. Black lines = excitatory projections, green lines = inhibitory projections, teal line = dopaminergic projection. ACC, anterior cingulate cortex; CD, caudate nucleus; GP, internal segment of the globus pallidus; HC, hippocampus; MDN, medial dorsal nucleus of the thalamus; PFC, prefrontal cortex; SN, substantia nigra pars compacta; VTA, ventral tegmental area.

cortical units, thereby forming a reverberating loop. At the same time, the PFC is part of a second excitatory reverberating loop through the medial dorsal (MD) nucleus of the thalamus (Alexander, DeLong, & Strick, 1986). These double reverberating loops maintain activation in the PFC working memory units during the hypothesis-testing procedure. However, the high spontaneous activity that is characteristic of the GABAergic neurons in the globus pallidus (GPi) tonically inhibit the thalamus, which prevents the closing of this cortical-thalamic loop, leading to the loss of information from working memory. To counteract this inhibition, the PFC excites medium spiny neurons (MSNs) in the head of the caudate nucleus (Bennett & Wilson, 2000), which in turn inhibit the pallidal neurons (since MSNs are GABAergic) that are inhibiting the thalamus. Reducing the pallidal inhibition of the thalamus allows reverberation in cortical-thalamic loops, and thereby facilitates working memory maintenance. A computational (i.e., spiking neuron) version of this model successfully accounts for many behavioral and single-neuron working memory-related phenomena (Ashby, Eil, Valentin, & Casale, 2005).

When feedback convinces a subject that the current categorization rule is incorrect, then a new rule must be selected and executive attention must be switched from the old rule to the new rule. In COVIS, these selection and switching operations are mediated by separate neural processes. The process of generating new candidate hypotheses is clearly complex, and a complete model of rule selection has not yet been formulated, in part because it likely involves poorly understood phenomena such as creativity and insight. In the computational version of the model, the anterior cingulate selects among alternative rules by enhancing the activity of the specific PFC working memory unit that represents a particular rule (Ashby, Valentin, & Turken, 2002). COVIS predicts that the most effortful and time-consuming processing follows error feedback, because negative feedback suggests that the current rule is incorrect. In support of this prediction, during the period of initial learning, slower response times on trials following an error predict higher accuracy in RB tasks, but not in II tasks (Tam, Maddox, & Huang-Pollock, 2013). When this process leads to the selection of the correct rule, there is often a sudden transition from suboptimal to optimal performance accompanied by an abrupt transition in neural firing of PFC neuron ensembles (Durstewitz, Vittoz, Floresco, & Seamans, 2010; Smith & Eil, 2015).

COVIS assumes there are three different classes of rules. The most privileged class includes the single rule that is currently active. This rule is maintained in a working memory loop and is the current focus of executive attention. The second most privileged class are those rules that can quickly become the focus of attention. These are candidate

rules each with their own reverberating working memory loops through PFC. The least privileged class is a much larger set of possible rules that are currently not in working memory and instead are represented in long-term memory. If properly activated, rules in this set can be moved into the intermediate set—that is, they could become represented by a PFC working memory unit and then maintained by their own reverberating circuit.

To perform well in RB tasks, participants must remember which rules they have already tested and rejected, in order to avoid revisiting these failed rules again. If the task is simple enough, then working memory might be sufficient for this task. Thus, COVIS predicts normal learning by medial temporal lobe amnesiacs in simple RB tasks in which the correct rule can be discovered before the list of rejected hypotheses is lost from working memory. In more difficult RB tasks (e.g., with many alternative rules), the search for the correct rule will exceed working memory capacity, so COVIS predicts that in these cases medial temporal lobe amnesiacs will be impaired. Much evidence supports the former prediction (Janowsky, Shimamura, Kritchevsky, & Squire, 1989; Leng & Parkin, 1988), but to our knowledge the latter prediction has not been rigorously tested. Even so, several studies have reported normal performance by amnesiacs on the first 50 trials of a difficult task, but impaired performance later on (Hopkins, Myers, Shohamy, Grossman, & Gluck, 2004; Knowlton et al., 1996). Thus, as in many other models, the projection from the PFC to the hippocampus is assumed to mediate the transition from working memory to long-term declarative memory representations (e.g., Eichenbaum & Cohen, 2001). Temporal cortex has also been shown to interact with PFC in rule knowledge retrieval from long-term storage (for a review see Bunge, 2004).

For successful RB performance, participants not only have to select candidate rules and maintain them in their working or long-term memory, but they also have to be able to give up on, or switch away from a rule that is no longer working. Switching, by definition, involves a change in the focus of executive attention, which has been long associated with the PFC. Patients with PFC lesions are not only impaired in working memory, but they also perseverate—i.e., they fail to switch rules after incorrect feedback (e.g., Kimberg, D'Esposito, & Farah, 1997). Recent work has suggested that norepinephrine plays a role in rule switching by influencing the PFC (Lapiz, Bondi, & Morilak, 2007; Lapiz & Morilak, 2006; McGaughy, Ross, & Eichenbaum, 2008; Tait et al., 2007). Accordingly, in COVIS, PFC plays a role in rule switching. The switch of attention away from a rule is initiated by a reduction in the PFC excitatory input to the head of the caudate nucleus (see Fig. 7.2). Such deactivation causes activation in the GPi to increase back

to its high baseline levels, which in turn increases inhibition on the thalamus, thereby breaking the cortical-thalamic loop. Therefore, the decision to switch attention originates in the PFC, but the switching itself is mediated within the basal ganglia (BG). This theory predicts that frontal patients perseverate because of decreased cortical control of the caudate nucleus. Neuroimaging and single-cell recording evidence supports this view that rule learning is mediated by interactions between the PFC and the BG (Badre & Frank, 2012; Bunge, 2004; Helie, Roeder, & Ashby, 2010; Pasupathy & Miller, 2005; Seger & Miller, 2010).

Evidence supporting the hypothesis that the BG play an important role in attentional switching comes from several sources. First, injections of a glutamate agonist directly into the striatum increases the frequency with which cats switch from one motor activity to another in a task where food rewards are delivered for such switching behaviors (Jaspers, De Vries, & Cools, 1990). Second, lesioning the dopamine (DA) fibers that project from the ventral tegmental area (VTA) into the PFC improves the performance of monkeys in an analogue of the Wisconsin Card Sorting Test (Roberts et al., 1994), which is an RB category learning task commonly used in neuropsychological assessment. If switching occurs in the PFC, then such lesions should impair switching performance. However, the PFC tonically inhibits the VTA (i.e., via a negative feedback loop). Lesioning the DA fibers projecting into the PFC releases the VTA from this inhibition. As a consequence, such lesions increase DA release into the BG (Roberts et al., 1994). If the BG are responsible for switching, and if switching is enhanced by DA, then lesioning DA fibers into PFC should improve switching, which is exactly the result observed by Roberts et al. (1994). Third, Racht-Delatour and El Massioui (1999) demonstrated that rats with lesions to the dorsal striatum had no deficits in learning which arm of a radial-arm maze was initially baited, but they did have deficits, relative to rats with sham lesions, when the position of the baited arm was successively switched according to a simple rule. Fourth, numerous studies have shown that Parkinson's disease patients, who have abnormally low levels of DA in the striatum, have a greater tendency to perseverate (Brown & Marsden, 1988), and are impaired in rule learning in a similar manner to patients with focal BG lesions (Ell, Marchant, & Ivry, 2006; Ell, Weinstein, & Ivry, 2010).

In conclusion, the COVIS declarative system includes multiple subprocesses such as selecting, focusing on, remembering, switching between, and increasing salience of successful rules while ignoring failed rules. Neuroimaging and neuropsychological results have provided evidence for such multiple, distinct processes in RB category learning (Kehagia, Cools, Barker, & Robbins, 2009; Monchi, Petrides,

Petre, Worsley, & Dagher, 2001; Price, Filoteo, & Maddox, 2009; Tachibana et al., 2009). Furthermore, DA influences these subprocesses (Ashby & Casale, 2003; Cools, 2006; Cools, Lewis, Clark, Barker, & Robbins, 2007; Frank & O'Reilly, 2006; Monchi et al., 2004; Moustafa & Gluck, 2011; Price et al., 2009; Seamans & Yang, 2004). Several recent computational modeling articles elegantly demonstrated the validity of the COVIS declarative system by showing that the model successfully accounts for RB category learning deficits that occur with normal aging, Parkinson's disease, or Anorexia Nervosa, and that it also accounts for the enhanced RB learning that occurs as a result of mild increases in positive mood—all by simply manipulating the amount of DA available to COVIS subprocesses (Filoteo et al., 2014; Hélie, Paul, & Ashby, 2012a, 2012b).

7.3.2 The COVIS Procedural-Learning System

The COVIS procedural-learning system incrementally learns arbitrary stimulus-response associations via dopamine-mediated reinforcement learning. Procedural-learning is typically associated with motor learning (e.g., Willingham, 1998; Willingham, Nissen, & Bullemer, 1989), and accordingly, the COVIS procedural system assumes that II learning includes a strong motor component.

Fig. 7.3 shows the architecture of the original 1998 COVIS procedural-learning system (Ashby et al., 1998; Ashby & Waldron, 1999). The key structure is the striatum, a major input region within the BG that includes the caudate nucleus and the putamen. In primates, all of the extrastriate visual cortex projects directly to the striatum, with about 10,000 visual cortical neurons converging on each striatal MSN (Wilson, 1995). The model assumes that, through a procedural-learning

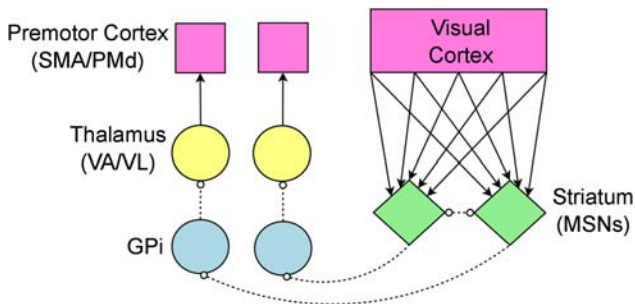


FIGURE 7.3 The 1998 version of the COVIS procedural category learning system. *GP_i*, internal segment of the globus pallidus; *MSN*, medium spiny neuron of the striatum; *SMA*, supplementary motor area; *VA/VL*, ventral anterior/ventral lateral nuclei of the thalamus.

process, each MSN associates an abstract motor program with a large group of visual cortical neurons (i.e., all that project to it). Much evidence supports the hypothesis that procedural-learning is mediated within the BG, and especially at cortical-striatal synapses, where synaptic plasticity is thought to follow reinforcement learning rules (Ashby & Ennis, 2006; Houk, Adams, & Barto, 1995; Mishkin, Malamut, & Bachevalier, 1984; Willingham, 1998). The COVIS procedural-learning system is a formal instantiation of these ideas.

The striatal MSNs send projections to a variety of prefrontal and premotor cortical areas. There are two synapses on this pathway. The first synapse is in the internal segment of the GPI, which is a major output structure within the BG. The second synapse is in the thalamus. Posterior regions of the putamen project primarily into motor and premotor areas of cortex [e.g., supplementary motor area (SMA)] via the ventral lateral nucleus of the thalamus (Matelli & Luppino, 1996). The SMA is densely interconnected with primary motor cortex and with other premotor areas (Dum & Strick, 2005). In contrast, the caudate and anterior putamen project to cortex primarily via the MD and ventral anterior (VA) thalamic nuclei. The MD nucleus projects widely into all anterior areas of frontal cortex, including PFC, whereas the primary cortical projection from VA is to preSMA and the supplementary eye fields (SEF) (Matelli & Luppino, 1996). SEF and preSMA are both densely interconnected with the PFC (Akkal, Dum, & Strick, 2007; Wang, Isoda, Matsuzaka, Shima, & Tanji, 2005), so the caudate projects primarily to prefrontal regions of cortex.

COVIS assumes that the procedural-learning in the striatum is facilitated by a DA-mediated reward signal from the substantia nigra pars compacta (SNpc). There is a large literature linking DA and reward, and many researchers have argued that a primary function of DA is to serve as the reward signal in reward-mediated learning (e.g., Houk et al., 1995; Wickens, 1993). The well-accepted theory is that positive feedback that follows successful behaviors increases phasic DA levels in the striatum, which has the effect of strengthening recently active synapses, whereas negative feedback causes DA levels to fall below baseline, which has the effect of weakening recently active synapses (e.g., Arbuthnott, Ingham, & Wickens, 2000; Calabresi, Pisani, Mercuri, & Bernardi, 1996; Reynolds & Wickens, 2002). In this way, the DA response to feedback serves as a teaching signal for which successful behaviors increase in probability and unsuccessful behaviors decrease in probability. According to this account, synaptic plasticity (long-term potentiation, LTP, or long-term depression, LTD) can only occur when the visual trace of the stimulus and the postsynaptic effects of DA overlap in time.

More specifically, synaptic plasticity in the striatum is strongest when the intracellular signaling cascades driven by *N*-methyl-D-aspartate (NMDA) receptor activation and DA D1 receptor activation coincide (Lisman, Schulman, & Cline, 2002; Rudy, 2014). The further apart in time these two cascades peak, the less effect DA will have on synaptic plasticity. For example, Yagishita et al. (2014) recently reported that synaptic plasticity was best (i.e., greatest increase in spine volume on striatal MSNs) when DA neurons were stimulated 600 ms after MSNs. When the DA neurons were stimulated before the MSNs or 5 s after the MSNs, then no evidence of any plasticity was observed. Similar results have been reported in II category learning. First, Worthy, Markman, and Maddox (2013) reported that II learning is best with feedback delays of 500 ms and slightly worse with delays of 0 or 1000 ms. Second, several studies have reported that feedback delays of 2.5 s or longer impair II learning, whereas delays as long as 10 s have no effect on RB category learning (Dunn, Newell, & Kalish, 2012; Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005). Valentin, Maddox, and Ashby (2014) showed that the COVIS procedural-learning system can accurately account for the effects of all these feedback delays.

There have been three different significant generalizations of the COVIS procedural-learning system during the past decade. One important extension was to add striatal cholinergic interneurons (known as TANs, for tonically active neurons) to the model (Ashby & Crossley, 2011; Crossley, Ashby, & Maddox, 2013, 2014; Crossley, Horvitz, Balsam, & Ashby, 2016). This generalization of the COVIS procedural-learning model is supported by a wide variety of neuroscience evidence (e.g., Apicella, Legallet, & Trouche, 1997; Matsumoto, Minamimoto, Graybiel, & Kimura, 2001; Pakhotin & Bracci, 2007; Smith, Raju, Pare, & Sidibe, 2004; for other evidence, see Ashby & Crossley, 2011). The revised architecture is shown in Fig. 7.4. The model assumes that the TANs tonically inhibit cortical input to striatal output neurons. The TANs are driven by neurons in the centromedian–parafascicular (CM-Pf) nuclei of the thalamus, which in turn are broadly tuned to features of the environment. In rewarding environments, the TANs learn to pause to stimuli that predict reward, which releases the cortical input to the striatum from inhibition. This allows striatal output neurons to respond to excitatory cortical input, thereby facilitating cortical-striatal plasticity. In this way, TAN pauses facilitate the learning and expression of striatal-dependent behaviors. When rewards are no longer available, the TANs cease to pause, which prevents striatal-dependent responding and protects striatal learning from decay.

Extending the COVIS procedural-learning system to include TANs allows the model to account for many new phenomena—some of which have posed difficult challenges for previous learning theories. One of

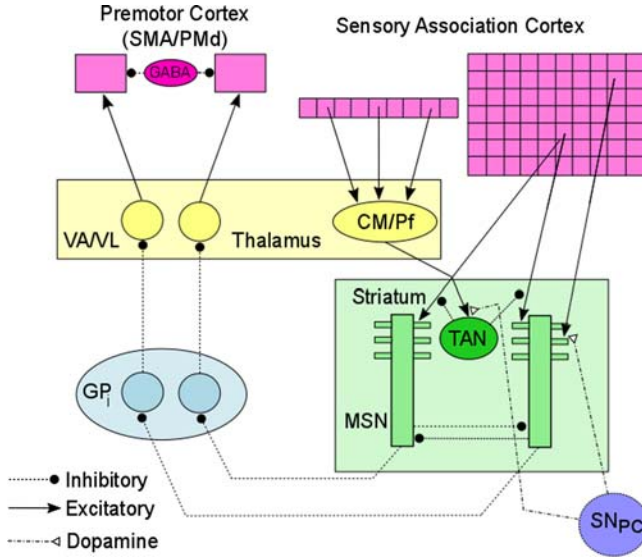


FIGURE 7.4 The architecture of the original version of the COVIS procedural category learning system with the addition of striatal cholinergic interneurons. *CM-Pf*, centromedian/parafascicular nuclei of the thalamus; *GP_i*, internal segment of the globus pallidus; *MSN*, medium spiny neuron of the striatum; *SMA*, supplementary motor area; *SN_{PC}*, substantia nigra pars compacta; *TAN*, tonically active interneuron of the striatum; *VAVL*, ventral anterior/ventral lateral nuclei of the thalamus.

these is that the reacquisition of an instrumental behavior after it has been extinguished is considerably faster than during original acquisition (Ashby & Crossley, 2011). The model accounts for this ubiquitous phenomenon because the withholding of rewards during the extinction period causes the TANs to stop pausing to sensory cues in the conditioning environment (since they are no longer associated with reward). This closes the gate between cortex and the striatum, which prevents further weakening of the cortical-striatal synapses. When the rewards are reintroduced, the TANs relearn to pause, and the behavior immediately reappears because of the preserved synaptic strengths. The Fig. 7.4 model also accounts for a related conditioning phenomenon known as renewal (Crossley et al., 2014), as well as many other instrumental conditioning phenomena (Crossley et al., 2016).

A second recent and significant extension of the COVIS procedural-learning model was to add a second stage of learning (Cantwell et al., 2015). This two-stage version of COVIS is illustrated in Fig. 7.5. Before this extension, virtually all theories of category learning, including COVIS, assumed that humans learn new categories by gradually forming associations directly between stimuli and responses. A strong

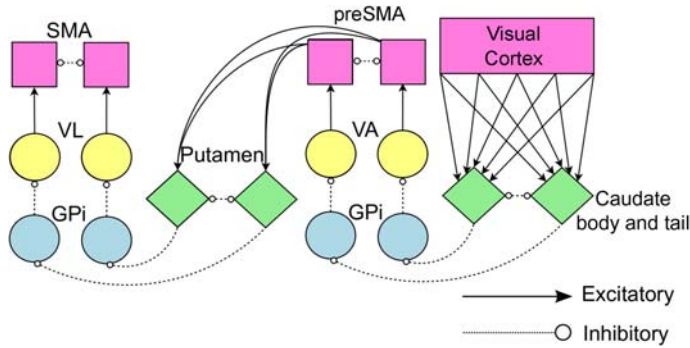


FIGURE 7.5 The architecture of the COVIS procedural category learning system that includes two stages of learning. *GP_i*, internal segment of the globus pallidus; *SMA*, supplementary motor area; *VA*, ventral anterior nucleus of the thalamus; *VL*, ventral lateral nucleus of the thalamus.

prediction of any such theory is that reversing the correct responses for all stimuli should cause catastrophic interference because recovery from a full reversal would require unlearning all prior stimulus-response associations, followed by new learning of the reversed associations. In contrast, creating new categories from the same stimuli in any other way should be less disruptive, because only some of the associations would have to be relearned, but not all. A number of studies, however, have reported that reversal learning is easier than learning novel categories (Cantwell et al., 2015; Kruschke, 1996; Maddox, Glass, O'Brien, Filoteo, & Ashby, 2010; Sanders, 1971; Wills, Noury, Moberly, & Newport, 2006). These results suggest that II category learning includes at least two stages: one that learns about category structure and another that learns about response mappings. With novel categories, new learning is required at both stages. In contrast, in the case of a full reversal, the category structures remain unchanged—only the response mappings must be relearned. The Fig. 7.5 model assumes that category structure is learned at synapses between visual cortex and the striatum. Cantwell et al. (2015) reported that recovery from a full reversal is impaired if the feedback is delayed by a few seconds, which suggests that the second stage is likely also in the BG. To account for this result, the Fig. 7.5 model, which synthesizes evidence of multiple distinct cortical-striatal loops into a neurocomputational theory, situates the second stage at synapses between preSMA and the putamen.

The third significant extension of the COVIS procedural-learning system during the past decade allows the model to account for how behaviors that are learned procedurally can eventually come to be executed automatically. Ashby et al. (2007) proposed that in contrast to early procedural-learning, which depends critically on the striatum,

automatic II categorization is mediated entirely within cortex and that the development of II automaticity is associated with a gradual transfer of control from the striatum to cortical-cortical projections from the relevant sensory areas directly to the premotor areas that initiate the behavior. According to this account, a critical function of the BG is to train purely cortical representations of automatic behaviors. The idea is that, via DA-mediated reinforcement learning, the BG learn to activate the correct postsynaptic targets in premotor cortex, which allows the appropriate cortical-cortical synapses to be strengthened via Hebbian learning. Once the cortical-cortical synapses have been built, the BG are no longer required to produce the automatic behavior. The learning rules in cortex and the striatum were hypothesized to be different because the low levels of cortical DA active transporter (DAT) prevent the rapid fluctuations in cortical DA levels needed for true reinforcement learning. In contrast, the striatum is rich in DAT, so there, synaptic plasticity is thought to follow reinforcement learning rules.

This generalization of the COVIS procedural system accounts for behavioral changes that occur as automaticity develops (i.e., improvements in both accuracy and response time), but it also accounts for a variety of neuroscience results that are problematic for other theories of automaticity (Ashby et al., 2007). For example, it correctly predicts that inactivation of the GPi (which essentially prevents the BG from influencing the cortical motor and premotor areas) does not disrupt the ability of monkeys to fluidly produce an over-learned motor sequence (Desmurget & Turner, 2010), and that Parkinson's patients, who have significant striatal dysfunction and are impaired in some RB and II tasks (see the next chapter), are relatively normal in executing automatic behaviors (Asmus, Huber, Gasser, & Schöls, 2008).

7.3.2.1 Interactions Between the COVIS Declarative and Procedural Systems

If human category learning is mediated by multiple systems, then an important question is to ask how the various systems interact and how their separate contributions are coordinated during the process of response selection. Much of the available evidence suggests that procedural and declarative memory systems compete for control of motor resources. For example, neuroimaging studies have found an antagonistic relationship between hippocampal and striatal activity (Dagher, Owen, Boecker, & Brooks, 2001; Jenkins, Brooks, Nixon, Frackowiak, & Passingham, 1994; Moody, Bookheimer, Vanek, & Knowlton, 2004; Nomura et al., 2007; Poldrack & Gabrieli, 2001; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999; Poldrack, Wagner, et al., 1999), and animal lesion studies have found that medial temporal lobe lesions can improve performance in striatal-mediated learning tasks, whereas

striatal lesions can improve performance in medial temporal lobe dependent tasks (Mitchell & Hall, 1988; O'Keefe & Nadel, 1978). However, more recent neuroimaging studies (Foerde, Knowlton, & Poldrack, 2006) have reported persistent striatal activation even during declarative control, and Crossley and Ashby (2015) recently reported that procedural category learning occurs even during declarative control. Both of these results suggest that these two systems learn simultaneously.

On the other hand, none of these results address the question of whether trial-by-trial switching between the systems is possible. All current category learning theories, including COVIS, predict that trial-by-trial switching should be effortless (Ashby et al., 1998; Erickson & Kruschke, 1998). Even so, we know of only two studies that address this issue (Ashby & Crossley, 2010; Erickson, 2008). Both studies used experiments that required participants to switch between procedural and declarative categorization strategies on a trial-by-trial basis in order to achieve optimal performance. Ashby and Crossley (2010) reported that only 2 of 53 participants showed any evidence of trial-by-trial switching, whereas Erickson (2008) reported that when more switching cues were provided, 51 of 123 participants (about 40%) successfully switched between systems on a trial-by-trial basis. Thus, the available behavioral data suggests that trial-by-trial switching between systems is much more difficult than originally thought.

Ashby and Crossley (2010) speculated that system switching might be mediated via the hyperdirect pathway through the BG. The hyperdirect pathway begins with direct excitatory projections from frontal cortex (e.g., preSMA) to the subthalamic nucleus, which is an input structure within the BG. The subthalamic nucleus then sends excitatory projections to the GPi (Joel & Weiner, 1997; Parent & Hazrati, 1995), which tends to offset inhibitory input from the striatum, making it more difficult for striatal activity to affect cortex. Hence, the hyperdirect pathway could permit (by reducing subthalamic activity) or prevent (by increasing subthalamic activity) striatal outputs from influencing cortex.

Evidence in support of this model comes from studies using the stop signal task, in which participants are required to initiate a motor response as quickly as possible when a cue is presented. On some trials, however, a second cue is presented soon after the first and in these cases participants are required to inhibit their response. A variety of evidence implicates the subthalamic nucleus in this task (Aron, Behrens, Smith, Frank, & Poldrack, 2007; Aron & Poldrack, 2006; Mostofsky & Simmonds, 2008). A popular model is that the second cue generates a stop signal in cortex that is rapidly transmitted through the hyperdirect pathway to the GPi, where it cancels out the go signal being sent through the striatum. Ashby and Crossley (2010) hypothesized

that when declarative memory is controlling behavior a similar stop signal may be used to inhibit a potentially competing response signal generated by the procedural memory system.

7.3.3 Dissociations Between RB and II Categorization

7.3.3.1 During Early Learning

During the past 15 years or so, many different experiments have tested parameter-free a priori predictions made by COVIS. These results are briefly summarized in Tables 7.1–7.3. Table 7.1 lists experimental manipulations that affect II learning and/or performance more than RB, Table 7.2 lists manipulations that affect RB learning and/or performance more than II, and Table 7.3 lists manipulations that affect both

TABLE 7.1 Some Empirical Dissociations That Affect Information-Integration Category Learning More Than Rule-Based Category Learning

Manipulation	Citations
Unsupervised learning	Ashby, Queller, and Berretty (1999) Ell, Ashby, and Hutchinson (2012)
Observational training	Ashby, Maddox, and Bohil (2002) Dunn et al. (2012)
Delayed feedback	Maddox et al. (2003) Maddox and Ing (2005)
Deferred (batch) feedback	Smith et al. (2014) Ashby, Ell, and Waldron (2003)
Response mapping	Maddox, Bohil, and Ing (2004) Maddox, Lauritzen, and Ing (2007) Spiering and Ashby (2008b)
Category label/location	Maddox et al. (2010)
Analogical transfer	Maddox, Filoteo, Lauritzen, Connally, and Hejl (2005) Casale, Roeder, and Ashby (2012)
Transfer to same-different	Helie and Ashby (2012)
Category discontinuity	Maddox et al. (2007) Maddox and Filoteo (2011)
Sequencing of training difficulty	Spiering and Ashby (2008a)
Category separation	Ell and Ashby (2006)

TABLE 7.2 Some Empirical Dissociations That Affect Rule-Based Category Learning More Than Information-Integration Category Learning

Manipulation	Citations
Concurrent working memory task	Waldron and Ashby (2001) Zeithamova and Maddox (2006)
Stress	Ell, Cosley and McCoy (2011)
Sleep deprivation	Maddox et al. (2009)
Feedback processing interference	Maddox et al. (2004) Filoteo, Lauritzen, and Maddox (2010) Zeithamova and Maddox (2007)
Positive mood	Nadler, Rabi, and Minda (2010)
Self regulation depletion	Minda and Rabi (2015)
Category number	Maddox, Filoteo, and Hejl (2004)

TABLE 7.3 Some Empirical Dissociations That Affect Rule-Based and Information-Integration Category Learning, But in Different (Often Opposite) Ways

Manipulation	Citations
Pressure	DeCaro et al. (2011) Markman, Maddox, and Worthy (2006)
Working memory capacity	DaCaro, Thomas, and Beilock (2008)
Richness of feedback	Maddox, Love, Glass, and Filoteo (2008)
Within versus across modality	Maddox, Ing, and Lauritzen (2006) Smith et al. (2015)
Feedback valence	Ashby and O'Brien (2007)
Dimensional priming	Grimm and Maddox (2013)
Visual masking	Hélie and Cousineau (2015)

tasks, but in different ways. Collectively, these results also provide strong evidence that learning in these tasks is mediated by separate systems.

The first four rows of [Table 7.1](#) summarize tests of COVIS predictions about how the nature and timing of trial-by-trial feedback affects response accuracy. In particular, COVIS predicts that, because the declarative system has access to working memory and executive

attention, it should be relatively unaffected by changes in the timing and form of the feedback signal. In contrast, the COVIS procedural system is highly sensitive to feedback parameters. First, because COVIS predicts that a DA-mediated reward signal is necessary for learning (e.g., LTP) to occur in the striatum, the absence or delay of such a reward signal should greatly interfere with this form of category learning. In support of these predictions (1) people can learn some RB categories in the absence of any trial-by-trial feedback, but there is no evidence that they can learn II categories (Table 7.1, Unsupervised learning), (2) II, but not RB learning is worse with observational training than with feedback training (Table 7.1, Observational training), (3) II learning is impaired with feedback delays as short as 2.5 s, whereas RB learning is unaffected by feedback delays as long as 10 s (Table 7.1, Delayed feedback), and (4) II learning is impaired if the feedback is batch delivered after every six trials, whereas RB learning is not (Table 7.1, Deferred feedback).

COVIS assumes that II learning is mediated by procedural memory, whereas RB learning is mediated by declarative memory (and especially working memory). One hallmark of procedural-learning is that it includes a motor component. For example, performance in procedural-learning-mediated tasks is impaired if the response locations are interchanged after learning has occurred (Willingham, Wells, Farrell, & Stemwedel, 2000). Rows 5 and 6 of Table 7.1 describe tests of this COVIS prediction. For example, as predicted, switching the response locations after learning impairs II performance but not RB performance.

Rows 7 and 8 of Table 7.1 report results showing that the knowledge acquired during II learning is specific to the stimulus-response associations practiced during training, whereas the knowledge acquired during RB learning is general and easily transferred to novel uses.

The first four rows of Table 7.2 describe results showing that RB learning is more impaired than II learning by (1) a simultaneous dual-task that recruits executive attention and working memory, (2) stress, (3) sleep deprivation, and (4) a reduction in the time available to process the feedback. These results, by themselves, are highly problematic for any proposal that RB versus II learning and performance differences are due to differences in task difficulty or complexity. As mentioned earlier, healthy young human adults learn the Fig. 7.1 RB categories much more quickly than the Fig. 7.1 II categories. Therefore, if the two category structures are learned by the same system, then any manipulation that impairs learning should have a greater effect in the more difficult II task than in the easier RB task. The first four rows of Table 7.2 describe exactly the opposite results. Row five of Table 7.2 describes the result that mild increases in positive mood facilitate RB learning, but not II learning.

Table 7.3 describes manipulations that affect both RB and II category learning, but in different ways. For example, DeCaro, Thomas, and Beilock (2008) reported that participants with a higher working memory span learned RB categories faster than participants with a lower span, but the lower span participants were actually better at learning II categories than the higher span participants. COVIS predicts that this interaction occurs because a high working memory span facilitates rule discovery, but it might also encourage a longer search for a logical rule in II conditions, which would slow learning, relative to participants who give up on declarative strategies earlier.

7.3.3.2 Automatic RB and II Responding are not Dissociable

Tables 7.1–7.3 establish many qualitative differences between initial RB and II learning and performance. In contrast, the best available evidence suggests that many of these differences disappear after the behaviors have been practiced long enough to become automatic (Ashby & Crossley, 2012; Hélie, Ell, & Ashby, 2015). For example, in a series of recent studies by Ashby and colleagues, participants were trained either on an RB or II task for almost 14,000 trials each (distributed over 23 sessions), with 4 of these sessions conducted at different stages of learning during fMRI scanning (Hélie, Waldschmidt, & Ashby, 2010; Hélie et al., 2010; Waldschmidt & Ashby, 2011). By the end of training, the accuracy and response time data from the different conditions were indistinguishable. Furthermore, although switching the location of the response keys interferes with II but not RB categorization early in learning (see Table 7.1), after automaticity had developed switching the response keys interfered equally with RB and II categorization, and to such an extent that there was almost no recovery from this interference over the course of 600 trials (Hélie et al., 2010). Similarly, although a dual-task interferes with initial RB but not II learning (see Table 7.2), after extensive training this difference also disappeared. Automatic RB and II categorization were both equally insensitive to dual-task interference (Hélie et al., 2010).

The neuroimaging results also showed convergence. In early training sessions, activation patterns for RB and II tasks were qualitatively different, but by session 20, many of these differences had disappeared (Soto, Waldschmidt, Hélie, & Ashby, 2013). For example, early RB performance was correlated with activation in PFC, the hippocampus, and the head of the caudate nucleus (Hélie et al., 2010), whereas early II training depended heavily on the putamen (Waldschmidt & Ashby, 2011). By session 20 however, activation in all of these areas no longer correlated with performance. Instead, only cortical activation (e.g., in premotor cortex) was positively correlated with response accuracy in both RB and II tasks. In the case of II tasks, these results are predicted

by the extension of the COVIS procedural system to automaticity described above (Ashby et al., 2007). This theory is currently evolving into a more general theory that accounts for the development of automaticity in both RB and II tasks. Essentially, the idea is that automatic RB and II categorization behaviors are mediated entirely within cortex by direct projections from the relevant sensory areas directly to the premotor areas that initiate the behavior (Helie & Ashby, 2009; Hélie et al., 2015). This theory predicts that although initial RB and II learning are mediated by very different neural systems that learn in qualitatively different ways, both systems have similar goals—namely, to train automatic cortical-cortical representations.

This more general theory of automaticity is still under development and so must be considered speculative. Laboratory experiments that study how automaticity develops are expensive to run, in both money and time. As a result, much less data is available to constrain theory and model building. Even so, there is some impressive existing evidence in support of this general theory. Some of this was described earlier (in the section entitled “The COVIS Procedural-Learning System”). Here we mention two other results related to the development of automaticity in RB tasks. First, several single-unit recording studies have reported that after training monkeys for months in a simple RB task, many rule-sensitive neurons can be found in PFC and in premotor cortex (Muhammad, Wallis, & Miller, 2006; Wallis & Miller, 2003). Importantly however, after such extensive training, the premotor rule-sensitive neurons respond in the task *before* the PFC rule-sensitive neurons (by 100 or more ms), suggesting that the automatic rule-guided behavior is no longer controlled by the PFC. Second, similar results have been reported for learning arbitrary stimulus-response associations—i.e., the PFC seems to play a critical role in initial learning, but not in the expression of well-learned associations (Puig & Miller, 2012). Understanding the role of the category learning systems in establishing representations of automatic categorization behaviors is likely to be an important focus of future research.

7.4 CONCLUSIONS

The preceding section demonstrates that COVIS provides a powerful description of the differences between the learning that dominates in RB and II category learning tasks. Even more impressive, however, is the fact that without COVIS, many of the experiments described in Tables 7.1–7.3 would never have been run. For example, it is difficult to imagine how any purely cognitive theory could ever predict that delaying feedback by a few seconds should interfere with one type of

category learning but not with another, or that switching the location of the response keys should cause a similar selective interference. Of course, one of the greatest benefits of a good theory is that it suggests new experiments to run, which themselves add to our knowledge in new and unexpected ways. In this sense, at least, COVIS has been extremely successful.

This chapter has focused on cognitive behavioral tests and predictions of COVIS. But the theory also makes a variety of other types of predictions—mostly because of its neurobiological underpinnings. In particular, it makes predictions about which neuropsychological patient groups should be impaired in category learning, and which groups should have normal category learning abilities. Further, when it predicts a deficit, it makes fairly specific predictions about the precise nature of these hypothesized deficits. The next chapter reviews the existing neuropsychological tests of COVIS, with a focus on patients with striatal or hippocampal dysfunction.

Despite its many successes, there are some recent signs that COVIS may be incomplete as a theory of human category learning. One important body of evidence pointing in this direction comes from a third type of category learning task called the “(A, not A) prototype distortion task” (Ashby & Maddox, 2005), in which each exemplar of one category is created by randomly distorting a single prototype (e.g., Posner & Keele, 1968). The subject’s task is to respond “Yes” or “No” depending on whether or not the presented stimulus is a member of this category (“not A” stimuli are just random patterns). Several studies have shown that a variety of neuropsychological patient groups that are known to have deficits in RB and/or II tasks show apparently normal (A, not A) prototype distortion learning, a result that is obviously problematic for the original version of COVIS. This list includes patients with Parkinson’s disease (Reber & Squire, 1999), schizophrenia (Keri, Kelemen, Benedek, & Janka, 2001), or Alzheimer’s disease (Sinha, 1999; although see Kéri et al., 1999).

COVIS hypothesizes that RB category learning uses working memory and other declarative memory systems (e.g., episodic and semantic memory), and that II learning uses procedural memory. One possibility is that performance in (A, not A) prototype distortion tasks is mediated by some different memory system. An obvious possibility is the perceptual representation memory system (Ashby & Casale, 2003; Reber & Squire, 1999). In support of this hypothesis, all of the existing neuroimaging studies of (A, not A) prototype distortion tasks have reported learning-related changes in occipital cortex (Aizenstein et al., 2000; Reber, Stark, & Squire, 1998a, 1998b; Zeithamova, Maddox, & Schnyer, 2008), and Casale and Ashby (2008) reported behavioral evidence that supports this position.

Learning is, by definition, the process of laying down some sort of memory trace, and there is certainly no reason to suspect that any of the separate memory systems that have been hypothesized are incapable of storing memories about categories. For this reason, a complete theory of human category learning is likely to assign some role to each of the different memory systems that have been identified by memory researchers.

Acknowledgments

Preparation of this chapter was supported by Public Health Service Grant MH2R01-063760.

References

- Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., Nebes, R. D., Larson, J. K., Ursu, S., & Carter, C. S. (2000). Complementary category learning systems identified using event-related functional mri. *Journal of Cognitive Neuroscience*, *12*(6), 977–987.
- Akkal, D., Dum, R. P., & Strick, P. L. (2007). Supplementary motor area and presupplementary motor area: targets of basal ganglia and cerebellar output. *The Journal of Neuroscience*, *27*(40), 10659–10673.
- 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*(1), 357–381.
- Apicella, P., Legallet, E., & Trouche, E. (1997). Responses of tonically discharging neurons in the monkey striatum to primary rewards delivered during different behavioral states. *Experimental Brain Research*, *116*(3), 456–466.
- Arbuthnott, G., Ingham, C., & Wickens, J. (2000). Dopamine and synaptic plasticity in the neostriatum. *Journal of Anatomy*, *196*(04), 587–596.
- Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., & Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (mri) and functional mri. *The Journal of Neuroscience*, *27*(14), 3743–3752.
- Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *The Journal of Neuroscience*, *26*(9), 2424–2433.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481.
- Ashby, F. G., & Casale, M. B. (2003). The cognitive neuroscience of implicit category learning. In L. Jiménez (Ed.), *Attention and implicit learning* (Vol. 48, pp. 109–142). John Benjamins Publishing Company.
- Ashby, F. G., & Crossley, M. J. (2010). Interactions between declarative and procedural-learning categorization systems. *Neurobiology of Learning and Memory*, *94*(1), 1–12.
- Ashby, F. G., & Crossley, M. J. (2011). A computational model of how cholinergic interneurons protect striatal-dependent learning. *Journal of Cognitive Neuroscience*, *23*(6), 1549–1566.
- Ashby, F. G., & Crossley, M. J. (2012). Automaticity and multiple memory systems. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(3), 363–376.

- 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*(11), 1728–1743.
- Ashby, F. G., Ell, S. W., & Waldron, E. M. (2003). Procedural learning in perceptual categorization. *Memory & Cognition*, *31*(7), 1114–1125.
- Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. *Psychology of Learning and Motivation*, *46*, 1–36.
- Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, *114*(3), 632–656.
- 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. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.
- Ashby, F. G., & Maddox, W. T. (2010). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147–161.
- 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., & O'Brien, J. B. (2007). The effects of positive versus negative feedback on information-integration category learning. *Perception & Psychophysics*, *69*, 865–878.
- Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, *61*(6), 1178–1199.
- Ashby, F. G., Valentin, V. V., & Turken, A. U. (2002). *Emotional cognition: From brain to behaviour*. Amsterdam: John Benjamins Publishing Company, chapter The effects of positive affect and arousal on working memory and executive attention: Neurobiology and computational models.
- Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review*, *6*(3), 363–378.
- Asmus, F., Huber, H., Gasser, T., & Schöls, L. (2008). Kick and rush paradoxical kinesia in parkinson disease. *Neurology*, *71*(9), 695.
- Badre, D., & Frank, M. J. (2012). Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: Evidence from fmri. *Cerebral Cortex*, *22*(3), 527–536.
- Bennett, B., & Wilson, C. (2000). Synaptology and physiology of neostriatal neurones. *Brain Dynamics and the Striatal Complex*, 111–140.
- Brown, R. G., & Marsden, C. D. (1988). Internal versus external cues and the control of attention in parkinson's disease. *Brain*, *111*(2), 323–345.
- Bunge, S. A. (2004). How we use rules to select actions: a review of evidence from cognitive neuroscience. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(4), 564–579.
- Calabresi, P., Pisani, A., Mercuri, N. B., & Bernardi, G. (1996). The corticostriatal projection: from synaptic plasticity to dysfunctions of the basal ganglia. *Trends in Neurosciences*, *19*(1), 19–24.
- Cantwell, G., Crossley, M. J., & Ashby, F. G. (2015). Multiple stages of learning in perceptual categorization: Evidence and neurocomputational theory. *Psychonomic Bulletin & Review*, *22*(6), 1598–1613.
- Casale, M. B., & Ashby, F. G. (2008). A role for the perceptual representation memory system in category learning. *Perception & Psychophysics*, *70*(6), 983–999.
- Casale, M. B., Roeder, J. L., & Ashby, F. G. (2012). Analogical transfer in perceptual categorization. *Memory & Cognition*, *40*(3), 434–449.
- Cools, R. (2006). Dopaminergic modulation of cognitive function—implications for l-dopa treatment in parkinson's disease. *Neuroscience & Biobehavioral Reviews*, *30*(1), 1–23.

- Cools, R., Lewis, S. J., Clark, L., Barker, R. A., & Robbins, T. W. (2007). L-dopa disrupts activity in the nucleus accumbens during reversal learning in parkinson's disease. *Neuropsychopharmacology*, 32(1), 180–189.
- Crossley, M. J., & Ashby, F. G. (2015). Procedural learning during declarative control. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 41(5), 1388–1403.
- Crossley, M. J., Ashby, F. G., & Maddox, W. T. (2013). Erasing the engram: The unlearning of procedural skills. *Journal of Experimental Psychology: General*, 142(3), 710–741.
- Crossley, M. J., Ashby, F. G., & Maddox, W. T. (2014). Context-dependent savings in procedural category learning. *Brain & Cognition*, 92, 1–10.
- Crossley, M. J., Horvitz, J. C., Balsam, P. D., & Ashby, F. G. (2016). Expanding the role of striatal cholinergic interneurons and the midbrain dopamine system in appetitive instrumental conditioning. *Journal of Neurophysiology*, 115, 240–254.
- Dagher, A., Owen, A. M., Boecker, H., & Brooks, D. J. (2001). The role of the striatum and hippocampus in planning. *Brain*, 124(5), 1020–1032.
- DeCaro, M. S., Thomas, R. D., Albert, N. B., & Beilock, S. L. (2011). Choking under pressure: multiple routes to skill failure. *Journal of Experimental Psychology: General*, 140(3), 390–406.
- 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(1), 284–294.
- Desmurget, M., & Turner, R. S. (2010). Motor sequences and the basal ganglia: kinematics, not habits. *The Journal of Neuroscience*, 30(22), 7685–7690.
- Dum, R. P., & Strick, P. L. (2005). Frontal lobe inputs to the digit representations of the motor areas on the lateral surface of the hemisphere. *The Journal of Neuroscience*, 25(6), 1375–1386.
- Dunn, J. C., Newell, B. R., & Kalish, M. L. (2012). The effect of feedback delay and feedback type on perceptual category learning: The limits of multiple systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 840–859.
- Durstewitz, D., Vitztoz, N. M., Floresco, S. B., & Seamans, J. K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron*, 66(3), 438–448.
- Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. Oxford University Press.
- Ell, S. W., & Ashby, F. G. (2006). The effects of category overlap on information–integration and rule–based category learning. *Perception & Psychophysics*, 68(6), 1013–1026.
- Ell, S. W., Ashby, F. G., & Hutchinson, S. (2012). Unsupervised category learning with integral-dimension stimuli. *The Quarterly Journal of Experimental Psychology*, 65(8), 1537–1562.
- Ell, S. W., Cosley, B., & McCoy, S. K. (2011). When bad stress goes good: increased threat reactivity predicts improved category learning performance. *Psychonomic Bulletin & Review*, 18(1), 96–102.
- Ell, S. W., Marchant, N. L., & Ivry, R. B. (2006). Focal putamen lesions impair learning in rule–based, but not information-integration categorization tasks. *Neuropsychologia*, 44(10), 1737–1751.
- Ell, S. W., Weinstein, A., & Ivry, R. B. (2010). Rule–based categorization deficits in focal basal ganglia lesion and parkinson's disease patients. *Neuropsychologia*, 48(10), 2974–2986.
- Erickson, M. A. (2008). Executive attention and task switching in category learning: Evidence for stimulus–dependent representation. *Memory & Cognition*, 36(4), 749–761.
- Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127(2), 107–140.

- Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science, 21*(3), 415–423.
- Filoteo, J. V., Maddox, W. T., Salmon, D. P., & Song, D. D. (2005). Information-integration category learning in patients with striatal dysfunction. *Neuropsychology, 19*(2), 212–222.
- Filoteo, J. V., Paul, E. J., Ashby, F. G., Frank, G. K., Helie, S., Rockwell, R., . . . Kaye, W. H. (2014). Simulating category learning and set shifting deficits in patients weight-restored from anorexia nervosa. *Neuropsychology, 28*(5), 741.
- Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences, 103*(31), 11778–11783.
- Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience, 120*(3), 497.
- Grimm, L. R., & Maddox, W. T. (2013). Differential impact of relevant and irrelevant dimension primes on rule-based and information-integration category learning. *Acta Psychologica, 144*(3), 530–537.
- Helie, S., & Ashby, F. G. (2009). A neurocomputational model of automaticity and maintenance of abstract rules. In *Neural networks, 2009. ijcn 2009. International joint conference on* (pp. 1192–1198).
- Helie, S., & Ashby, F. G. (2012). Learning and transfer of category knowledge in an indirect categorization task. *Psychological Research, 76*(3), 292–303.
- Hélie, S., & Cousineau, D. (2015). Differential effect of visual masking in perceptual categorization. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 816–825.
- Hélie, S., Ell, S. W., & Ashby, F. G. (2015). Learning robust cortico-cortical associations with the basal ganglia: An integrative review. *Cortex, 64*, 123–135.
- Hélie, S., Paul, E. J., & Ashby, F. G. (2012a). A neurocomputational account of cognitive deficits in parkinson's disease. *Neuropsychologia, 50*(9), 2290–2302.
- Hélie, S., Paul, E. J., & Ashby, F. G. (2012b). Simulating the effects of dopamine imbalance on cognition: From positive affect to parkinsons disease. *Neural Networks, 32*, 74–85.
- Helie, S., Roeder, J. L., & Ashby, F. G. (2010). Evidence for cortical automaticity in rule-based categorization. *The Journal of Neuroscience, 30*(42), 14225–14234.
- Hélie, S., Waldschmidt, J. G., & Ashby, F. G. (2010). Automaticity in rule-based and information-integration categorization. *Attention, Perception, & Psychophysics, 72*(4), 1013–1031.
- Hopkins, R. O., Myers, C. E., Shohamy, D., Grossman, S., & Gluck, M. (2004). Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia, 42*(4), 524–535.
- 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.
- 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*(3), 548–560.
- Jaspers, R., De Vries, T., & Cools, A. (1990). Enhancement in switching motor patterns following local application of the glutamate agonist ampa into the cat caudate nucleus. *Behavioural Brain Research, 37*(3), 237–246.
- Jenkins, I., Brooks, D., Nixon, P., Frackowiak, R., & Passingham, R. (1994). Motor sequence learning: a study with positron emission tomography. *The Journal of Neuroscience, 14*(6), 3775–3790.

- Joel, D., & Weiner, I. (1997). The connections of the primate subthalamic nucleus: indirect pathways and the open–interconnected scheme of basal ganglia–thalamocortical circuitry. *Brain Research Reviews*, 23(1), 62–78.
- Kehagia, A. A., Cools, R., Barker, R. A., & Robbins, T. W. (2009). Switching between abstract rules reflects disease severity but not dopaminergic status in parkinson's disease. *Neuropsychologia*, 47(4), 1117–1127.
- Kéri, S., Kálmán, J., Rapcsak, S. Z., Antal, A., Benedek, G., & Janka, Z. (1999). Classification learning in alzheimer's disease. *Brain*, 122(6), 1063–1068.
- Keri, S., Kelemen, O., Benedek, G., & Janka, Z. (2001). Intact prototype learning in schizophrenia. *Schizophrenia Research*, 52(3), 261–264.
- Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1997). Cognitive functions in the prefrontal cortex: Working memory and executive control. *Current Directions in Psychological Science*, 185–192.
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, 273(5280), 1399–1402.
- Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, 8(2), 225–247.
- Lapiz, M. D. S., Bondi, C. O., & Morilak, D. A. (2007). Chronic treatment with desipramine improves cognitive performance of rats in an attentional set-shifting test. *Neuropsychopharmacology*, 32(5), 1000–1010.
- Lapiz, M. D. S., & Morilak, D. A. (2006). Noradrenergic modulation of cognitive function in rat medial prefrontal cortex as measured by attentional set shifting capability. *Neuroscience*, 137(3), 1039–1049.
- Leng, N. R., & Parkin, A. J. (1988). Double dissociation of frontal dysfunction in organic amnesia. *British Journal of Clinical Psychology*, 27(4), 359–362.
- Lisman, J., Schulman, H., & Cline, H. (2002). The molecular basis of camkii function in synaptic and behavioural memory. *Nature Reviews Neuroscience*, 3(3), 175–190.
- 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., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule–based but not information–integration category learning. *Memory & Cognition*, 32(4), 582–591.
- Maddox, W. T., Bohil, C. J., & Ing, A. D. (2004). Evidence for a procedural-learning-based system in perceptual category learning. *Psychonomic Bulletin & Review*, 11(5), 945–952.
- Maddox, W. T., & Filoteo, J. V. (2011). Stimulus range and discontinuity effects on information-integration category learning and generalization. *Attention, Perception, & Psychophysics*, 73(4), 1279–1295.
- Maddox, W. T., Filoteo, J. V., Hejl, K. D., et al. (2004). Category number impacts rule-based but not information-integration category learning: further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 227–235.
- Maddox, W. T., Filoteo, J. V., Lauritzen, J. S., Connally, E., & Hejl, K. D. (2005). Discontinuous categories affect information-integration but not rule–based category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4), 654–669.
- Maddox, W. T., Glass, B. D., O'Brien, J. B., Filoteo, J. V., & Ashby, F. G. (2010). Category label and response location shifts in category learning. *Psychological Research*, 74(2), 219–236.
- Maddox, W. T., Glass, B. D., Wolosin, S. M., Savarie, Z. R., Bowen, C., Matthews, M. D., & Schnyer, D. M. (2009). The effects of sleep deprivation on information–integration categorization performance. *Sleep*, 32(11), 1439–1448.

- Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 100–107.
- Maddox, W. T., Ing, A. D., & Lauritzen, J. S. (2006). Stimulus modality interacts with category structure in perceptual category learning. *Perception & Psychophysics*, 68(7), 1176–1190.
- Maddox, W. T., Lauritzen, J. S., & Ing, A. D. (2007). Cognitive complexity effects in perceptual classification are dissociable. *Memory & Cognition*, 35(5), 885–894.
- Maddox, W. T., Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition*, 108(2), 578–589.
- Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science*, 17(11), 944–948.
- Matelli, M., & Luppino, G. (1996). Thalamic input to mesial and superior area 6 in the macaque monkey. *Journal of Comparative Neurology*, 372(1), 59–87.
- Matsumoto, N., Minamimoto, T., Graybiel, A. M., & Kimura, M. (2001). Neurons in the thalamic cm-pf complex supply striatal neurons with information about behaviorally significant sensory events. *Journal of Neurophysiology*, 85(2), 960–976.
- McCaughy, J., Ross, R., & Eichenbaum, H. (2008). Noradrenergic, but not cholinergic, deafferentation of prefrontal cortex impairs attentional set-shifting. *Neuroscience*, 153(1), 63–71.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24(1), 167–202.
- Minda, J. P., & Rabi, R. (2015). Ego depletion interferes with rule-defined category learning but not non-rule-defined category learning. *Frontiers in Psychology*, 6.
- 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, NY: Guilford Press.
- Mitchell, J. A., & Hall, G. (1988). Caudate-putamen lesions in the rat may impair or potentiate maze learning depending upon availability of stimulus cues and relevance of response cues. *The Quarterly Journal of Experimental Psychology*, 40(3), 243–258.
- Monchi, O., Petrides, M., Doyon, J., Postuma, R. B., Worsley, K., & Dagher, A. (2004). Neural bases of set-shifting deficits in parkinson's disease. *The Journal of Neuroscience*, 24(3), 702–710.
- 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. *The Journal of Neuroscience*, 21(19), 7733–7741.
- Moody, T. D., Bookheimer, S. Y., Vanek, Z., & Knowlton, B. J. (2004). An implicit learning task activates medial temporal lobe in patients with parkinson's disease. *Behavioral Neuroscience*, 118(2), 438–442.
- Mostofsky, S. H., & Simmonds, D. J. (2008). Response inhibition and response selection: two sides of the same coin. *Journal of Cognitive Neuroscience*, 20(5), 751–761.
- Moustafa, A. A., & Gluck, M. A. (2011). A neurocomputational model of dopamine and prefrontal-striatal interactions during multicue category learning by parkinson patients. *Journal of Cognitive Neuroscience*, 23(1), 151–167.
- 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(6), 974–989.
- Nadler, R. T., Rabi, R. R., & Minda, J. P. (2010). Better mood and better performance: Learning rule-described categories is enhanced by positive moods. *Psychological Science*, 21, 1770–1776.
- Nomura, E., Maddox, W., Filoteo, J., Ing, A., Gitelman, D., Parrish, T., ... Reber, P. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, 17(1), 37–43.

- Nosofsky, R. M., Stanton, R. D., & Zaki, S. R. (2005). Procedural interference in perceptual classification: Implicit learning or cognitive complexity? *Memory & Cognition*, 33(7), 1256–1271.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map* (Vol. 3). Oxford: Clarendon Press.
- Pakhotin, P., & Bracci, E. (2007). Cholinergic interneurons control the excitatory input to the striatum. *The Journal of Neuroscience*, 27(2), 391–400.
- Parent, A., & Hazrati, L.-N. (1995). Functional anatomy of the basal ganglia. ii. the place of subthalamic nucleus and external pallidum in basal ganglia circuitry. *Brain Research Reviews*, 20(1), 128–154.
- Pasupathy, A., & Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature*, 433(7028), 873–876.
- Poldrack, R. A., Clark, J., Pare-Blagoev, E., Shohamy, D., Moyano, J. C., Myers, C., & Gluck, M. (2001). Interactive memory systems in the human brain. *Nature*, 414(6863), 546–550.
- Poldrack, R. A., & Gabrieli, J. D. (2001). Characterizing the neural mechanisms of skill learning and repetition priming evidence from mirror reading. *Brain*, 124(1), 67–82.
- Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia*, 41(3), 245–251.
- Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology*, 13(4), 564–574.
- Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage*, 10(1), 15–35.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77(3p1), 353–363.
- Price, A., Filoteo, J. V., & Maddox, W. T. (2009). Rule-based category learning in patients with parkinson's disease. *Neuropsychologia*, 47(5), 1213–1226.
- Puig, M. V., & Miller, E. K. (2012). The role of prefrontal dopamine d1 receptors in the neural mechanisms of associative learning. *Neuron*, 74(5), 874–886.
- Racht-Delatour, B. V. G., & El Massioui, N. (1999). Rule-based learning impairment in rats with lesions to the dorsal striatum. *Neurobiology of Learning and Memory*, 72(1), 47–61.
- Reber, P. J., & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with parkinson's disease. *Behavioral Neuroscience*, 113(2), 235–242.
- Reber, P. J., Stark, C. E., & Squire, L. R. (1998a). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory*, 5(6), 420–428.
- Reber, P. J., Stark, C. E., & Squire, L. R. (1998b). Cortical areas supporting category learning identified using functional mri. *Proceedings of the National Academy of Sciences*, 95(2), 747–750.
- Reynolds, J. N., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, 15(4), 507–521.
- Roberts, A. C., De Salvia, M., Wilkinson, L., Collins, P., Muir, J., Everitt, B., & Robbins, T. (1994). 6-hydroxydopamine lesions of the prefrontal cortex in monkeys enhance performance on an analog of the wisconsin card sort test: possible interactions with sub-cortical dopamine. *The Journal of Neuroscience*, 14(5), 2531–2544.
- Rudy, J. W. (2014). *The Neurobiology of Learning and Memory*. Sunderland, MA: Sinauer.
- Sanders, B. (1971). Factors affecting reversal and nonreversal shifts in rats and children. *Journal of Comparative and Physiological Psychology*, 74, 192–202.
- Seamans, J. K., & Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in Neurobiology*, 74(1), 1–58.
- Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, 33, 203–219.
- Sinha, R. R. (1999). *Neuropsychological Substrates of Category Learning (Unpublished doctoral dissertation)*. Brown University.

- Smith, J. D., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., & Grace, R. C. (2011). Pigeons categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review*, *18*(2), 414–421.
- Smith, J. D., Berg, M. E., Cook, R. G., Murphy, M. S., Crossley, M. J., Boomer, J., ... others (2012). Implicit and explicit categorization: A tale of four species. *Neuroscience & Biobehavioral Reviews*, *36*(10), 2355–2369.
- Smith, J. D., Boomer, J., Zakrzewski, A. C., Roeder, J. L., Church, B. A., & Ashby, F. G. (2014). Deferred feedback sharply dissociates implicit and explicit category learning. *Psychological Science*, *25*, 447–457.
- Smith, J. D., & Ell, S. W. (2015). One giant leap for categorizers: One small step for categorization theory. *PLoS One*, *10*(9), e0137334.
- Smith, J. D., Johnston, J. J. R., Musgrave, R. D., Zakrzewski, A. C., Boomer, J., Church, B. A., & Ashby, F. G. (2015). Cross-modal information integration in category learning. *Attention, Perception, & Psychophysics*, in press.
- Smith, Y., Raju, D. V., Pare, J. -F., & Sidibe, M. (2004). The thalamostriatal system: A highly specific network of the basal ganglia circuitry. *Trends in Neurosciences*, *27*(9), 520–527.
- Soto, F. A., Waldschmidt, J. G., Helie, S., & Ashby, F. G. (2013). Brain activity across the development of automatic categorization: A comparison of categorization tasks using multi-voxel pattern analysis. *NeuroImage*, *71*, 284–297.
- Spiering, B. J., & Ashby, F. G. (2008a). Initial training with difficult items facilitates information-integration but not rule-based category learning. *Psychological Science*, *19* (11), 1169–1177.
- Spiering, B. J., & Ashby, F. G. (2008b). Response processes in information-integration category learning. *Neurobiology of Learning and Memory*, *90*(2), 330–338.
- Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171–177.
- Tachibana, K., Suzuki, K., Mori, E., Miura, N., Kawashima, R., Horie, K., ... Mushiaki, H. (2009). Neural activity in the human brain signals logical rule identification. *Journal of Neurophysiology*, *102*(3), 1526–1537.
- Tait, D. S., Brown, V. J., Farovik, A., Theobald, D. E., Dalley, J. W., & Robbins, T. W. (2007). Lesions of the dorsal noradrenergic bundle impair attentional set-shifting in the rat. *European Journal of Neuroscience*, *25*(12), 3719–3724.
- Tam, H., Maddox, W. T., & Huang-Pollock, C. L. (2013). Posterror slowing predicts rule-based but not information-integration category learning. *Psychonomic Bulletin & Review*, *20*(6), 1343–1349.
- Valentin, V. V., Maddox, W. T., & Ashby, F. G. (2014). A computational model of the temporal dynamics of plasticity in procedural learning: Sensitivity to feedback timing. *Frontiers in Psychology*, *5*(643). Available from: <http://dx.doi.org/10.3389/fpsyg.2014.00643>.
- Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*(1), 168–176.
- Waldschmidt, J. G., & Ashby, F. G. (2011). Cortical and striatal contributions to automaticity in information-integration categorization. *NeuroImage*, *56*(3), 1791–1802.
- Wallis, J. D., & Miller, E. K. (2003). From rule to response: neuronal processes in the pre-motor and prefrontal cortex. *Journal of Neurophysiology*, *90*(3), 1790–1806.
- Wang, Y., Isoda, M., Matsuzaka, Y., Shima, K., & Tanji, J. (2005). Prefrontal cortical cells projecting to the supplementary eye field and presupplementary motor area in the monkey. *Neuroscience Research*, *53*(1), 1–7.
- Wickens, J. (1993). *A theory of the striatum*. Pergamon Press.
- Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological Review*, *105*, 558–584.

- Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1047–1060.
- Willingham, D. B., Wells, L. A., Farrell, J. M., & Stemwedel, M. E. (2000). Implicit motor sequence learning is represented in response locations. *Memory & Cognition*, 28(3), 366–375.
- Wills, A., Noury, M., Moberly, N. J., & Newport, M. (2006). Formation of category representations. *Memory & Cognition*, 34(1), 17–27.
- 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: MIT Press.
- Worthy, D. A., Markman, A. B., & Maddox, W. T. (2013). Feedback and stimulus–offset timing effects in perceptual category learning. *Brain and Cognition*, 81(2), 283–293.
- Yagishita, S., Hayashi-Takagi, A., Ellis-Davies, G. C., Urakubo, H., Ishii, S., & Kasai, H. (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science*, 345(6204), 1616–1620.
- Zeithamova, D., & Maddox, W. T. (2006). Dual–task interference in perceptual category learning. *Memory & Cognition*, 34(2), 387–398.
- Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition*, 35, 1380–1398.
- Zeithamova, D., Maddox, W. T., & Schnyer, D. M. (2008). Dissociable prototype learning systems: evidence from brain imaging and behavior. *The Journal of Neuroscience*, 28(49), 13194–13201.