Category Learning in the Brain

Carol A. Seger[1] and **Earl K. Miller**[2]

[1]Department of Psychology and Program in Molecular, Cellular, and Integrative Neurosciences, Colorado State University, Fort Collins, Colorado 80523; Carol.Seger@colostate.edu

[2]The Picower Institute for Learning and Memory and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; ekmiller@mit.edu

Abstract

The ability to group items and events into functional categories is a fundamental characteristic of sophisticated thought. It is subserved by plasticity in many neural systems, including neocortical regions (sensory, prefrontal, parietal, and motor cortex), the medial temporal lobe, the basal ganglia, and midbrain dopaminergic systems. These systems interact during category learning. Corticostriatal loops may mediate recursive, bootstrapping interactions between fast reward-gated plasticity in the basal ganglia and slow reward-shaded plasticity in the cortex. This can provide a balance between acquisition of details of experiences and generalization across them. Interactions between the corticostriatal loops can integrate perceptual, response, and feedback-related aspects of the task and mediate the shift from novice to skilled performance. The basal ganglia and medial temporal lobe interact competitively or cooperatively, depending on the demands of the learning task.

Keywords

classification; concept learning; memory systems

INTRODUCTION

Although our brains can store specific experiences, it is not always advantageous for us to be too literal. A brain limited to storing an independent record of each experience would require a prodigious amount of storage and bog us down with details. We have instead evolved the ability to detect the higher-level structure of experiences, the commonalities across them that allow us to group experiences into meaningful categories and concepts. This process imbues the world with meaning. We instantly recognize and respond appropriately to objects, situations, expressions, etc., even though we have never encountered those exact examples before. It stimulates proactive, goal-directed thought by allowing us to generalize about (to imagine) future situations that share fundamental elements with past experience. Imagine the mental cacophony without this ability. The world would lack any deeper meaning. Experiences would be fragmented and unrelated. Things would seem strange and unfamiliar if they differed even trivially from previous examples. This situation describes many of the cognitive characteristics of neuropsychiatric disorders such as autism.

Here, we review how categories are learned by the brain. We begin with a brief definition of categories and describe how category learning is studied. We argue that categorization is not dependent on any single neural system, but rather results from the recruitment of a variety of neural systems depending on task demands. We then describe the primary brain areas involved in categorization learning: the visual cortex, the prefrontal and parietal cortices, the basal ganglia, and the medial temporal lobe. This leads to a discussion and hypotheses about how neural systems interact during category acquisition, which focus on interactions within and between corticostriatal loops connecting cortex and basal ganglia and between the basal ganglia and the medial temporal lobe. We end by summarizing principles by which the brain learns categories and other abstractions.

Categories

Categories represent our knowledge of groupings and patterns that are not explicit in the bottom-up sensory inputs. A simple example is crickets sharply dividing a range of pure tones into mate versus bat (a predator) (Wyttenbach et al. 1996). A wide range of tones on either side of a sharp boundary (16 kHz) are treated equivalently, whereas nearby tones that straddle it are treated differently. This grouping of experience by functional relevance occurs at many levels of processing and for a wide range of phenomena from more literal (e.g., color) to abstract (e.g., peace, love, and understanding). Many categorical distinctions are innate or result from many years of experience (e.g., faces), but key to human intelligence is our ability to learn new categories quickly, even when they are multivariate and abstract (e.g., Free Jazz, gastropub). Mahon & Caramazza (2009) and Martin (2007), among others, have published excellent reviews of innate or well-learned categories. We focus on category learning.

Examples of category learning tasks are shown in Figure 1. Many tasks use novel stimuli formed according to a particular perceptual manipulation and then grouped according to an experimenter-defined boundary. Some examples of stimuli used in tasks include prototypes, information integration, and stimuli morphed along a continuum (Figure 1*a–c*, respectively). We can also group events and actions into categories by more abstract properties or rules, which can range from simple deterministic rules based on a single easily identified dimension to more complex situations in which rules may be probabilistic, or complex (e.g., a conjunctive or disjunctive rule), or require identification of an abstract feature not actually present in the physical item (e.g., the rule "same" or "different" (Figure 1*e*). Categorization can even be completely arbitrary (Figure 1*d*). For example, imagine a group of students, half of whom are enrolled in one section of a course and half in the other section. The students within each section likely do not share any particular perceptual characteristics that are not shared by students in the other section. However, this categorization scheme has great utility for predicting which students are likely to attend class in a particular room at a particular time.

BRAIN AREAS INVOLVED IN CATEGORY LEARNING

Not surprisingly given the variety of above examples, category learning likely involves many brain systems including most of the neocortex, the hippocampus, and the basal ganglia. We review which types of category tasks recruit each region of the brain and describe each region's putative role. We make no claims to an exhaustive treatment; categorical representations are likely in many domains and their respective neural systems. For example, evidence indicates that the amygdala participates in generalization of knowledge about fearful or aversive types of stimuli (Barot et al. 2008).

Visual Cortex

We focus on the visual system, the best-studied modality. However, similar processes are likely present in other sensory modalities, including the auditory (Vallabha et al. 2007), the somatosensory (Romo & Salinas 2001), and the olfactory (Howard et al. 2009) systems.

Likely candidates for visual categorization are areas at the highest levels of visual processing. One is the inferior temporal cortex (ITC), whose neurons have complex shape selectivity (Desimone et al. 1984, Logothetis & Sheinberg 1996, Tanaka 1996). Investigators have known about neurons with category-like tuning properties since the seminal work on "face cells" by Gross and colleagues (Desimone et al. 1984). The human fusiform face area (Kanwisher et al. 1997), an ITC area with a preponderance of face cells, is recruited during learning of new face categories (DeGutis & D'Esposito 2007). Inferior temporal neurons in trained monkeys are specifically activated by trees or fish and show relatively little differentiation within those categories (Vogels 1999). Microstimulation of monkey ITC can facilitate visual classification of novel images (Kawasaki & Sheinberg 2008).

However, the ITC may play less of a role in learning explicit representations of category membership and more of a role in high-level analysis of features that contribute to categorization. ITC neurons often do not completely generalize among category members; they retain selectivity for underlying perceptual similarity between individuals (Freedman et al. 2003, Jiang et al. 2007). They also emphasize certain critical stimuli or diagnostic features for the categories and show greater activity for stimuli near category boundaries (DeGutis & D'Esposito 2007, Freedman et al. 2003, Sigala & Logothetis 2002, Baker et al., 2002).

Simple shape-based perceptual categories may be acquired in earlier visual areas. A commonly used task is the dot pattern prototype learning task (see Figure 1*a*); subjects learn a single category (e.g., "A" versus "not A") via simply observing category members. This type of relatively simple category learning may depend on plasticity in the early visual system locus. fMRI studies show activity changes after dot pattern learning in the extrastriate visual cortex, typically around BA 18/19 and roughly corresponding to visual area V2 (Aizenstein et al. 2000; Reber et al. 1998, 2003). Performance on this task is preserved in persons with amnesia (Knowlton & Squire 1993; for a comprehensive review, see Smith 2008), indicating independence from the medial temporal lobe memory system, and is preserved in Parkinson disease (Reber & Squire 1999), indicating independence from corticostriatal systems. However, patients with moderate-severity Alzheimer disease, which can include damage to extrastriate visual cortex, are impaired (Keri et al. 1999). Other categorization tasks, however, do recruit corticostriatal and/or medial temporal lobe systems, especially more complex category learning that involves learning via trial-and-error feedback and learning of multiple categories (Casale & Ashby 2008, Little & Thulborn 2005, Vogels et al. 2002).

Plasticity in visual cortex likely involves local changes in the strength of cortical synapses owing to Hebbian learning (McClelland 2006) subserved by mechanisms of long-term potentiation. Sensory cortex typically emphasizes stability over plasticity, particularly in adults. Thus, perceptual categories, especially in early sensory cortex, do not usually result from just casual or limited amounts of passive experience with a stimulus.

Prefrontal Cortex

The prefrontal cortex occupies a far greater proportion of the human cerebral cortex than it does in other animals, which suggests that it might contribute to those cognitive capacities

that separate humans from animals (Fuster 1995, Miller & Cohen 2001). It seems more readily modifiable by experience than does the sensory cortex.

For example, Freedman and colleagues (2001, 2002, 2003) trained monkeys to categorize stimuli along a morphing continuum of different blends of "cats" and "dogs" (see Figure 1*c*) and found a large proportion of randomly selected lateral prefrontal cortex (PFC) neurons with hallmarks of category representations: sharp differences in activity to similar-looking stimuli across a discrete category boundary yet similar activity to different-looking members of the same category. Simultaneous recording from the PFC and anterior-ventral ITC revealed weaker category effects in the ITC (they retained more selectivity for individual members) and that category signals appeared with a shorter latency in the PFC than in the ITC, as if it were fed back from the PFC (Freedman et al. 2003, Meyers et al. 2008). Human imaging studies found that ITC is sensitive to perceptual features of stimuli and perceptual distance between stimuli, but only PFC represents the boundary between actual categories or crucial conjunctions between features (Jiang et al. 2007, Li et al. 2009).

PFC neurons also reflect abstract rule-based categorical distinctions. For example, Wallis and Miller (Muhammad et al. 2006, Wallis et al. 2001, Wallis & Miller 2003,) trained monkeys to apply either a "same" or "different" rule to novel pairs of pictures (see Figure 1*e*). Many PFC neurons conveyed which rule was in effect independent of which specific cue signaled the rule, was not linked to the behavioral response, and was unaffected by the exact pictures the monkeys were judging. By contrast, the rules had relatively little effect in the ITC, even though the ITC is directly connected with the lateral PFC and it is critical for visual analysis of the pictures (Muhammad et al. 2006).

Parietal Cortex

The parietal cortex seems to emphasize visuospatial functions and linking information from perceptual cortex with potential responses. Many studies have examined its neural selectivity by having subjects discriminate the direction of motion of moving dots. Many direction-selective neurons are in extrastriate area V5/MT (Newsome et al. 1986) and project to the lateral inferior parietal lobe and insula, which integrate overall movement pattern (Ho et al. 2009, Rorie & Newsome 2005). Freedman & Assad (2006) trained monkeys to classify 360° of motion direction into two categories and found that category membership was strongly reflected in the lateral inferior parietal region, but much less so in V5/MT. The respective roles of the parietal and frontal cortices in categorization and visual cognition in general remain to be determined, but several studies indicate a close functional link between the lateral inferior parietal lobe and the PFC (Buschman & Miller 2007, Chafee & Goldman-Rakic 2000).

Premotor and Motor Cortex

Categorical decision tasks also involve selection and execution of an appropriate behavior. This recruits premotor cortex (PMC) and primary motor cortex within the frontal lobe. Category learning can also result in plasticity in brain systems involved in attention and eye movements (Blair et al. 2009). Little & Thulborn (2005) found changes in frontal eye field and supplementary eye field activity across training in a dot pattern categorization task that likely reflected improved visual scanning of the stimuli.

As expertise is developed, reliance on motor systems increases and reliance on other systems decreases. Indeed, PFC damage preferentially affects new learning: Animals and humans can still engage in complex behaviors as long as they were well learned before the damage (Dias et al. 1997, Murray et al. 2000, Shallice 1982). PFC neurons are more strongly activated during new learning than during execution of familiar tasks (Asaad et al.

1998). There are stronger signals in the dorsal PMC than in the PFC when humans performed familiar versus novel classifications (Boettiger & D'Esposito 2005) and when monkeys performed familiar abstract rules (Muhammad et al. 2006). Thus, the PFC may acquire new categories, but other areas such as the PMC may execute them once they become familiar.

Hippocampus and the Medial Temporal Lobe

The medial temporal lobe (MTL) has anatomical and functional connections with cortex and seems specialized for rapid learning of individual instances (O'Reilly & Munakata 2000). The circuitry of the MTL and cortex forms a loop: Information from broad neocortical regions across the parietal, frontal, and temporal cortices projects to the entorhinal region of the parahippocampal gyrus. From the entorhinal cortex, the primary projections pass to the dentate gyrus, the CA3 field of the hippocampus, the CA1 field, and back to the entorhinal cortex. The CA3 field contains autoassociative recurrent links, which allow association formation during encoding and pattern completion during recall (Becker & Wojtowicz 2007, Gluck et al. 2003, O'Reilly & Munakata 2000).

Several lines of evidence suggest multiple roles for the MTL in categorization. Categorization can make use of the MTL's ability to learn individual instances. One task that requires instance learning is the arbitrary categorization task (Figure 1*d*), in which the category membership of each item must be remembered individually. fMRI studies find that MTL (among other systems, including corticostriatal systems) is often recruited during these tasks (Poldrack et al. 1999, 2001; Seger & Cincotta 2005). Likewise, monkey neurophysiology studies found that neurons in the hippocampus and temporal cortex show category-specific activity after training monkeys to group arbitrary stimuli (Hampson et al. 2004). Kreiman et al. (2000) found neurons in the human MTL that were selective for diverse pictures of familiar concepts such as Bill Clinton. The MTL's instance-learning capacity may also be invoked to store exceptions to rules and other categorical regularities (Love et al. 2004). Some degree of instance memory may be required in all categorization tasks that use novel stimuli; the MTL may be required to set up a memory representation of each stimulus that can then be accessed by other systems (Meeter et al. 2008).

Another important potential contribution of the MTL follows from observations that information acquired via the MTL can be transferred to new situations. One example is acquired equivalence. For example, if a subject learns that stimulus A is in categories 1 and 2, and stimulus B is in category 1, they can reasonably infer that stimulus B might also be in category 2. The MTL is involved in these tasks (Myers et al. 2003, Shohamy & Wagner 2008).

The Basal Ganglia and Corticostriatal Loops

The basal ganglia are a collection of subcortical nuclei that interact with cortex in corticostriatal loops. Cortical inputs arrive largely via the striatum and ultimately are directed back into the cortex via the thalamus. The basal ganglia maintain a degree of topographical separation in different loops, ensuring that the output is largely to the same cortical areas that gave rise to the initial inputs to the basal ganglia (Alexander et al. 1986, Hoover & Strick 1993, Kelly & Strick 2004, Parthasarathy et al. 1992). The frontal cortex receives the largest portion of BG outputs, suggesting some form of close collaboration between these structures (Middleton & Strick 1994, 2002). However, almost all cortical regions participate in corticostriatal loops (Flaherty & Graybiel 1991, Kemp & Powell 1970). Although there is overlap between the loops at their boundaries, it is useful to talk of four loops: executive, motivational, visual, and motor (Lawrence et al. 1998, Seger 2008), as illustrated in Figure 2. The basal ganglia exert a tonic inhibition on cortex; they selectively

and phasically release the cortex to allow for selection of a movement (Humphries et al. 2006) or cognitive strategy (Frank 2005). In categorization tasks, this function may be recruited to help with selection of both an appropriate category representation and related strategies or behaviors (Seger 2008).

The basal ganglia are active in a wide variety of categorization tasks (Nomura et al. 2007; Poldrack et al. 1999, 2001; Seger & Cincotta 2005; Zeithamova et al. 2008), particularly those that require subjects to learn via trial and error (Cincotta & Seger 2007, Merchant et al. 1997). Performance on these tasks is impaired in patients with compromised basal ganglia functions owing to Parkinson and Huntington disease (Ashby & Maddox 2005, Knowlton et al. 1996, Shohamy et al. 2004). The roles of individual corticostriatal loops and their interactions during categorization are discussed further below.

Midbrain Dopaminergic System and Reinforcement Learning Mechanisms

Any form of supervised (reward-based) learning, including category learning, depends on the midbrain dopaminergic brain systems (the ventral tegmental area and the substantia nigra, pars compacta) (Schultz et al. 1992). Neurons in these areas show activity that seems to correspond to the reward prediction error signals suggested by animal learning models (Hollerman & Schultz 1998, Montague et al. 2004; but see Redgrave & Gurney 2006). They activate and release dopamine widely throughout the basal ganglia and cortex (especially in the frontal lobe) whenever animals are unexpectedly rewarded, and they pause when an expected reward is withheld. Over time the cells learn to respond to an event that directly predicts a reward: The event stands in for the reward (Schultz et al. 1993). Functional imaging has found that the basal ganglia, a primary target of dopamine neurons, are also sensitive to prediction error (Seymour et al. 2007).

Cortical inputs converge onto the dendrites of striatal spiny cells along with a strong input from midbrain dopaminergic neurons. Dopamine is required for synapse strengthening or weakening in the striatum by long-term depression or potentiation, respectively (Calabresi et al. 1992, Kerr & Wickens 2001, Otani et al. 1998). These anatomical and neurophysiological properties suggest that the striatum has an ideal infrastructure for rapid, reward-gated, supervised learning that quickly forms representations of the patterns of cortical connections that predict reward (Houk & Wise 1995, Miller & Buschman 2007). Functional imaging, neuropsychological, and computational studies suggest that feedback-based category learning via trial and error depends on both dopamine and the basal ganglia (Shohamy et al. 2008).

INTERACTION BETWEEN NEURAL SYSTEMS DURING CATEGORY LEARNING

Above, we discussed how categorization learning relies on multiple neural systems. For example, a visual categorization task may recruit the visual cortex and the MTL to represent and memorize the individual stimuli and facilitate processing of relevant features, the prefrontal cortex to learn and represent categorization rules and strategies, and the basal ganglia, parietal lobe, and motor cortices to make decisions and select behavioral responses on the basis of categorical information. In this section we discuss several ways that these neural systems may interact during category learning.

Interactions Between Fast Subcortical Plasticity and Slower Cortical Plasticity

A key issue in learning is the need to balance the advantages and disadvantages of fast versus slow plasticity (see sidebar, Computational Factors in Category Learning). Fast plasticity (large changes in synaptic weights with each episode) in a neural network has

advantages in rapid storage of relevant activity patterns (and quick learning). But slow plasticity (small weight changes) allows networks to generalize; gradual changes result in neural ensembles that are not tied to specific inputs but instead store what is common among them. One possible solution is to have fast plasticity and slow plasticity systems interact (McClelland et al. 1995, O'Reilly & Munakata, 2000). For example, McClelland et al. (1995) suggested that long-term memory consolidation results from fast plasticity in the hippocampus, the output of which trains slower plasticity cortical networks that gradually elaborate the memories and link them to others. A similar relationship between the cerebellum and the cortex could underlie motor learning (Houk & Wise 1995). We suggest that an interaction between fast plasticity in the basal ganglia and slow plasticity in the cortex underlies many forms of category learning and abstraction (Miller & Buschman 2007).

Fast and slow plasticity may arise from different applications of the dopaminergic teaching signal. Both the cortex and the basal ganglia receive projections from midbrain dopaminergic neurons, but dopamine input to cortex is much lighter than that into the striatum (Lynd-Balta & Haber 1994). Dopamine projections also show a gradient in connectivity with heavier inputs in the PFC that drop off posteriorly (Goldman-Rakic et al. 1989, Thierry et al. 1973). This observation may explain why the PFC seems to show a greater deal of experience-dependent selectivity than does the visual cortex. In the striatum, the dopamine influence may be greater still. Dopamine neurons terminate near the synapse between a cortical axon and striatal spiny cell, a good position to gate plasticity between the cortex and the striatum. DA neurons synapse on the dendrites of cortical neurons, and therefore may have a lesser influence. Thus, whereas plasticity in the striatum may be fast and reward-gated in the cortex, it may be slower and reward-shaded. The striatum may be better suited to learn details, the specific cues, responses, etc. that predict rewards, whereas the cortex acquires the commonalities among them that result in categories and abstractions (see Daw et al. 2005).

Some evidence suggests this notion. Pasupathy & Miller (2005) found that during conditional visuomotor learning in monkeys, striatal neural activity showed rapid, almost bistable, changes compared with a much slower trend in the PFC. Seger & Cincotta (2006) found that as humans learn rules, changes in striatal activity precede those in the frontal cortex. Abstract rules are more strongly represented (more neurons and stronger effects) and appear with a shorter latency in the frontal cortex than in the dorsal striatum (Muhammad et al. 2006), which is consistent with a greater cortical involvement in abstraction.

Under this view, normal learning depends on balance between the fast and the slow plasticity systems. An imbalance between these systems that causes basal ganglia plasticity to become abnormally strong and overwhelm the cortex might result in an autistic-like brain that is overwhelmed with details and cannot generalize. Recent work by Bear and colleagues may provide a molecular link (Dolen et al. 2007). They found that many psychiatric and neurological symptoms of Fragile X, including autism, can be explained by abnormally high activation of metabotropic glutamate receptor mGluR5. MGluR5 colocalizes with dopamine receptors in striatal neurons and is thought to regulate dopamine-dependent plasticity. The idea is that too much mGluR5 boosts dopaminergic plasticity mechanisms in striatum and overwhelms the cortex, resulting in an inability to generalize and fractionated, piecemeal cognition.

Interactions Within Corticostriatal Loops: Recursive Processing and Bootstrapping

As noted above, the cortex forms closed anatomical loops with the basal ganglia: Channels within the basal ganglia return outputs, via the thalamus, to the same cortical areas that gave rise to their initial cortical input (Hoover & Strick 1993, Kelly & Strick 2004). Closed loops

suggest recursivity, bootstrapping operations in which the results from one iteration are fed back through the loop for further processing and elaboration. Some form of recursive processing must underlie the open-ended nature of human memory and thought. We suggest that recursive interactions between basal ganglia fast plasticity and slow cortical plasticity underlie construction of categories and abstractions. This idea may be reflected in a hallmark of human intelligence: It is easiest for us to understand new categories and concepts if they can be grounded first in familiar ones. We learn to multiply through serial addition, and we understand quantum mechanics by constructing analogies to waves and particles.

Interactions Between Corticostriatal Loops

Although basal gangliay–PFC connections are particularly prominent, the basal ganglia interact with all cortical regions. Figure 2 illustrates the major patterns of projection, broken into loops. Functional imaging has shown that all four loops are recruited during categorization learning, albeit in different roles (Seger 2008, Seger & Cincotta 2005). The visual loop receives information from visual cortex; this information feeds forward to the executive and motor loops, providing a potential mechanism for selection of appropriate responses (Ashby et al. 1998, 2007), as well as back to visual cortex where it may assist in refinement of visual processing. The executive loop is associated with functions necessary for categorization learning, including feedback processing, working memory updating, and set shifting. The motor loop is involved in selecting and executing appropriate motor behavior, including selection of the motor response used to indicate category membership. The motivational loop is involved in processing reward and feedback.

The loops interact during learning. Seger and colleagues (2010) examined interactions between corticostriatal loops during categorization using Granger causality modeling and found patterns consistent with directed influence from the visual loop to the motor loop, and from the motor loop to the executive loop. This pattern is consistent with the processes required during each step of a typical categorization trial: processing the visual stimulus, preparing and executing the motor response indicating category membership, and receiving and processing feedback.

Corticostriatal loops also interact across many experiences or trials as subjects progress from being novices to experts in a categorization domain. The executive and motivational loops are most important early, when acquisition of information is fastest and feedback processing is the most useful, whereas the motor loop rises in importance as expertise is acquired (Williams & Eskandar 2006). The anterior caudate (executive loop) is sensitive to learning rate; activity is greatest when learning is occurring most rapidly (Williams & Eskandar 2006) and there is the greatest amount of prediction error (difference between expected outcome and actual outcome) to serve as a learning signal (Haruno & Kawato 2006). In contrast, the putamen (motor loop) is more engaged late in learning, when the category membership (and associated reward or feedback) is well learned. (Seger et al. 2010, Williams & Eskandar 2006). This idea is consistent with observations that the rodent dorsomedial striatum (equivalent to primate anterior caudate) is important for initial goal-oriented learning, whereas dorsolateral striatum (equivalent to primate posterior putamen) is important for later habit formation (Yin & Knowlton 2006).

Finally, corticostriatal loops can compete depending on the material being learned. Categories that can be learned via explicit rule-based processes tend to rely on the PFC and anterior caudate regions involved in the executive loop. Other category structures (such as information integration categories; Figure 1*b*) that are learned via more implicit processes rely on the visual loop. The COVIS model (Ashby et al. 1998, 2007) proposes that the executive and visual loops compete for dominance in controlling categorization. This

proposal is supported by studies examining individual differences in prefrontal capacity: Subjects with high capacity tend to favor the rule-learning system and are relatively impaired at learning an information integration task that requires the more implicit strategy to achieve optimal performance (Decaro et al. 2008).

Interactions Between the Medial Temporal Lobe and Basal Ganglia

Both MTL and BG systems can form relationships between stimuli and categories. As described above, the MTL does so via explicit representation of the stimulus and its arbitrary category membership, whereas the basal ganglia map perceptual commonalities of categories to their associated behaviors. Human imaging studies suggest competition between MTL and BG systems during category learning: As BG activity increases, MTL activity decreases (Poldrack et al. 1999, 2001). However, relative decreases in MTL activity are difficult to interpret in functional imaging studies; apparent suppression of the MTL may simply be due to lower activity during categorization than during the comparison tasks (Law et al. 2005). Stronger evidence for competition between the two systems comes from lesion and pharmacological manipulations. When MTL is damaged or inhibited, the BG can take over a larger role in the control of behavior (Frank et al. 2006). Subjects with basal ganglia damage due to Parkinson disease recruit MTL to a larger extent than do controls during probabilistic classification category learning (Moody et al. 2004).

The BG and MTL may not invariably compete during categorization learning. Some studies show parallel recruitment of both systems, implying independent or cooperative contributions (Cincotta & Seger 2007). The MTL may be required initially to set up new individual item representations of stimuli (Meeter et al. 2008). These stimulus representations may then be accessible to BG systems for forming associations between stimuli and categories. Consistent with this theory, Poldrack et al. (2001) found transient MTL activity at the beginning of a probabilistic classification task, which was then followed by a relative decrease in MTL activity and increase in BG activity.

It is unclear how interaction between MTL and BG is mediated. Some evidence indicates that the relationship is bilateral: Increases in BG activity lead to decreases in MTL activity and vice versa (Lee et al. 2008). Some research indicates that PFC is involved in this process (Poldrack & Rodriguez 2004). During distraction with PFC demanding dual tasks, categorization performance becomes more strongly related to striatal activity and less related to MTL (Foerde et al. 2006). In emotional situations, the amygdala can likely mediate the balance between systems (Wingard & Packard 2008).

CONCLUSION: PRINCIPLES OF CATEGORY LEARNING IN THE BRAIN

We are only beginning to understand how the brain learns categories. But we can posit some potential principles and hypotheses.

■   Categorization involves both stimulus representations (e.g., of features, central tendencies, and degree of variability) and processes (e.g., decision-making processes establishing a criterion or rule for category membership) that recruit different neural systems depending on the type of category and how it is used.

■   The brain does not have one single "categorization area." Categories are represented in a distributed fashion across the brain, and multiple neural systems are involved. Many of the systems involved in categorization have been identified in the multiple memory systems framework (Ashby & O'Brien 2005, Poldrack & Foerde 2008, Smith & Grossman 2008). Categorization tasks are not process-pure: Multiple systems may be recruited to solve any given categorization problem.

- ■ Category learning withstands fundamental computational constraints. A trade-off exists between generalizing across previous experience and remembering specific items and events. This trade-off may be solved by having fast plasticity (large synaptic weight changes) in subcortical systems (e.g., basal ganglia and hippocampus) train slower plasticity (smaller weight changes) in the cortex, the latter of which builds the category representations by finding the commonalities across the specifics learned by the former. Normal learning depends on balance between these mechanisms. The balance can change depending on task demands. Certain neuropsychiatric disorders, such as autism, may result from an imbalance that causes the faster plasticity mechanisms in the subcortex to overwhelm the slower cortical plasticity, which could result in a brain that has great difficulty generalizing.

- ■ Category learning may depend on recursive, bootstrapping interactions within corticostriatal loops. The open-ended nature of human thought likely depends on some form of recursive processing, and the closed anatomical loops the basal ganglia form with the cortex seem well suited. Different phases of learning and different aspects of a categorization task may also involve interactions across different corticostriatal loops.

- ■ Category learning cuts across distinctions between implicit and explicit systems and declarative and nondeclarative memory systems. Explicit systems are those that are associated with some degree of conscious penetrability (Seger 1994). In categorization, these include PFC systems recruited in explicit rule-learning tasks, as well as MTL systems that result in consciously accessible episodic memories. Most other systems are typically considered to be implicit or unconscious (e.g., perceptual cortex); however, some (notably the corticostriatal loops) can be recruited in both explicit and implicit tasks. The declarative–nondeclarative distinction differs from the explicit–implicit distinction because it separates MTL-dependent memory processes (declarative) from other learning systems (nondeclarative). Categorization tasks may recruit various combinations of implicit and/or explicit, declarative and/or nondeclarative systems. For example, simple dot pattern learning is largely implicit (it occurs without intention to learn or awareness of learning) and nondeclarative (it is independent of MTL systems). Rule learning is explicit because subjects intend to learn and have awareness of what they have learned and is nondeclarative because it largely recruits prefrontal cortex and does not require the MTL for acquisition.

- ■ A major challenge in understanding category learning is determining which category-learning systems are recruited in particular situations, and whether the systems function independently, cooperatively, or antagonistically. What is ultimately learned is an interaction between the structure of the information in the environment and the neural systems recruited to process the information (Reber et al. 2003, Zeithamova et al. 2008). Which systems are recruited can also depend on factors that can vary across individuals and situations, such as cognitive capacity (Decaro et al. 2008) and motivational state (Grimm et al. 2007).

Acknowledgments

Glossary

ITC	inferotemporal cortex
PFC	prefrontal cortex
MTL	medial temporal lobe

LITERATURE CITED

Aizenstein HJ, MacDonald AW, Stenger VA, Nebes RD, Larson JK, et al. Complementary category learning systems identified using event-related functional MRI. J. Cogn. Neurosci. 2000; 12:977–87. [PubMed: 11177418]

Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu. Rev. Neurosci. 1986; 9:357–81. [PubMed: 3085570]

Asaad WF, Rainer G, Miller EK. Neural activity in the primate prefrontal cortex during associative learning. Neuron. 1998; 21:1399–407. [PubMed: 9883732]

Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. A neuropsychological theory of multiple systems in category learning. Psychol. Rev. 1998; 105:442–81. [PubMed: 9697427]

Ashby FG, Ennis JM, Spiering BJ. A neurobiological theory of automaticity in perceptual categorization. Psychol. Rev. 2007; 114:632–56. [PubMed: 17638499]

Ashby FG, Maddox WT. Human category learning. Annu. Rev. Psychol. 2005; 56:149–78. [PubMed: 15709932]

Ashby FG, O'Brien JB. Category learning and multiple memory systems. Trends Cogn. Sci. 2005; 9:83–89. [PubMed: 15668101]

Baker CI, Behrmann M, Olson CR. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nat. Neurosci. 2002; 5:1210–16. [PubMed: 12379864]

Barot SK, Kyono Y, Clark EW, Bernstein IL. Visualizing stimulus convergence in amygdala neurons during associative learning. Proc. Natl. Acad. Sci. USA. 2008; 105:20959–63. [PubMed: 19091953]

Becker S, Wojtowicz JM. A model of hippocampal neurogenesis in memory and mood disorders. Trends Cogn. Sci. 2007; 11:70–76. [PubMed: 17174137]

Blair MR, Watson MR, Walshe RC, Maj F. Extremely selective attention: eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. J. Exp. Psychol. Learn. Mem. Cogn. 2009; 35:1196–206. [PubMed: 19686015]

Boettiger CA, D'Esposito M. Frontal networks for learning and executing arbitrary stimulus-response associations. J. Neurosci. 2005; 25:2723–32. [PubMed: 15758182]

Buschman TJ, Miller EK. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science. 2007; 315:1860–62. [PubMed: 17395832]

Calabresi P, Maj R, Pisani A, Mercuri NB, Bernardi G. Long-term synaptic depression in the striatum: physiological and pharmacological characterization. J. Neurosci. 1992; 12:4224–33. [PubMed: 1359031]

Casale MB, Ashby FG. A role for the perceptual representation memory system in category learning. Percept. Psychophys. 2008; 70:983–99. [PubMed: 18717385]

Chafee MV, Goldman-Rakic PS. Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. J. Neurophysiol. 2000; 83:1550–66. [PubMed: 10712479]

Cincotta CM, Seger CA. Dissociation between striatal regions while learning to categorize via feedback and via observation. J. Cogn. Neurosci. 2007; 19:249–65. [PubMed: 17280514]

Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. 2005; 8:1704–11. [PubMed: 16286932]

Decaro MS, Thomas RD, Beilock SL. Individual differences in category learning: sometimes less working memory capacity is better than more. Cognition. 2008; 107:284–94. [PubMed: 17707363]

DeGutis J, D'Esposito M. Distinct mechanisms in visual category learning. Cogn. Affect. Behav. Neurosci. 2007; 7:251–59. [PubMed: 17993211]

Desimone R, Albright TD, Gross CG, Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. J. Neurosci. 1984; 4:2051–62. [PubMed: 6470767]

Dias R, Robbins TW, Roberts AC. Dissociable forms of inhibitory control within prefrontal cortex with an analog of the Wisconsin Card Sort Test: restriction to novel situations and independence from "on-line" processing. J. Neurosci. 1997; 17:9285–97. [PubMed: 9364074]

Dolen G, Osterweil E, Rao BS, Smith GB, Auerbach BD, et al. Correction of fragile X syndrome in mice. Neuron. 2007; 56:955–62. [PubMed: 18093519]

Flaherty AW, Graybiel AM. Corticostriatal transformations in the primate somatosensory system. Projections from physiologically mapped body-part representations. J. Neurophysiol. 1991; 66:1249–63. [PubMed: 1722244]

Foerde K, Knowlton BJ, Poldrack RA. Modulation of competing memory systems by distraction. Proc. Natl. Acad. Sci. USA. 2006; 103:11778–83. [PubMed: 16868087]

Frank MJ. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. J. Cogn. Neurosci. 2005; 17:51–72. [PubMed: 15701239]

Frank MJ, O'Reilly RC, Curran T. When memory fails, intuition reigns: Midazolam enhances implicit inference in humans. Psychol. Sci. 2006; 17:700–7. [PubMed: 16913953]

Freedman DJ, Assad JA. Experience-dependent representation of visual categories in parietal cortex. Nature. 2006; 443:85–88. [PubMed: 16936716]

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representation of visual stimuli in the primate prefrontal cortex. Science. 2001; 291:312–16. [PubMed: 11209083]

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. J. Neurophysiol. 2002; 88:914–28. [PubMed: 12163541]

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J. Neurosci. 2003; 23:5235–46. [PubMed: 12832548]

Fuster, JM. Memory in the Cerebral Cortex. MIT Press; Cambridge, MA: 1995.

Gluck MA, Meeter M, Myers CE. Computational models of the hippocampal region: linking incremental learning and episodic memory. Trends Cogn. Sci. 2003; 7:269–76. [PubMed: 12804694]

Goldman-Rakic PS, Leranth C, Williams SM, Mons N, Geffard M. Dopamine synaptic complex with pyramidal neurons in primate cerebral cortex. Proc. Natl. Acad. Sci. USA. 1989; 86:9015–19. [PubMed: 2573073]

Grimm LR, Markman AB, Maddox WT, Baldwin GC. Differential effects of regulatory fit on category learning. J. Exp. Soc. Psychol. 2007; 44:920–27. [PubMed: 19412324]

Hampson RE, Pons TP, Stanford TR, Deadwyler SA. Categorization in the monkey hippocampus: a possible mechanism for encoding information into memory. Proc. Natl. Acad. Sci. USA. 2004; 101:3184–89. [PubMed: 14978264]

Haruno M, Kawato M. Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. J. Neurophysiol. 2006; 95:948–59. [PubMed: 16192338]

Ho TC, Brown S, Serences JT. Domain general mechanisms of perceptual decision making in human cortex. J. Neurosci. 2009; 29:8675–87. [PubMed: 19587274]

Hollerman JR, Schultz W. Dopamine neurons report an error in the temporal prediction of reward during learning. Nat. Neurosci. 1998; 1:304–9. [PubMed: 10195164]

Hoover JE, Strick PL. Multiple output channels in the basal ganglia. Science. 1993; 259:819–21. [PubMed: 7679223]

Houk JC, Wise SP. Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. Cereb. Cortex. 1995; 5:95–110. [PubMed: 7620294]

Howard JD, Plailly J, Grueschow M, Haynes JD, Gottfried JA. Odor quality coding and categorization in human posterior piriform cortex. Nat. Neurosci. 2009; 12(7):932–38. [PubMed: 19483688]

Humphries MD, Stewart RD, Gurney KN. A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. J. Neurosci. 2006; 26:12921–42. [PubMed: 17167083]

Jiang X, Bradley E, Rini RA, Zeffiro T, Vanmeter J, Riesenhuber M. Categorization training results in shape- and category-selective human neural plasticity. Neuron. 2007; 53:891–903. [PubMed: 17359923]

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 1997; 17:4302–11. [PubMed: 9151747]

Kawasaki K, Sheinberg DL. Learning to recognize visual objects with microstimulation in inferior temporal cortex. J. Neurophysiol. 2008; 100:197–211. [PubMed: 18463185]

Kelly RM, Strick PL. Macro-architecture of basal ganglia loops with the cerebral cortex: use of rabies virus to reveal multisynaptic circuits. Prog. Brain Res. 2004; 143:449–59. [PubMed: 14653187]

Kemp JM, Powell TP. The cortico-striate projection in the monkey. Brain. 1970; 93:525–46. [PubMed: 4990231]

Keri S, Kalman J, Rapcsak SZ, Antal A, Benedek G, Janka Z. Classification learning in Alzheimer's disease. Brain. 1999; 122:1063–68. [PubMed: 10356059]

Kerr JND, Wickens JR. Dopamine D-1/D-5 receptor activation is required for long-term potentiation in the rat neostriatum in vitro. J. Neurophysiol. 2001; 85:117–24. [PubMed: 11152712]

Knowlton BK, Mangels JA, Squire LR. A neostriatal habit learning system in humans. Science. 1996; 273:1399–402. [PubMed: 8703077]

Knowlton BK, Squire LR. The learning of categories: parallel brain systems for item memory and category knowledge. Science. 1993; 262:1747–49. [PubMed: 8259522]

Kreiman G, Koch C, Fried I. Category-specific visual responses of single neurons in the human medial temporal lobe. Nat. Neurosci. 2000; 3:946–53. [PubMed: 10966627]

Law JR, Flanery MA, Wirth S, Yanike M, Smith AC, et al. Functional magnetic resonance imaging activity during the gradual acquisition and expression of paired-associate memory. J. Neurosci. 2005; 25:5720–29. [PubMed: 15958738]

Lawrence AD, Sahakian BJ, Robbins TW. Cognitive functions and corticostriatal circuits: insights from Huntington's disease. Trends Cogn. Sci. 1998; 2:379–88. [PubMed: 21227253]

Lee AS, Duman RS, Pittenger C. A double dissociation revealing bidirectional competition between striatum and hippocampus during learning. Proc. Natl. Acad. Sci. USA. 2008; 105:17163–68. [PubMed: 18955704]

Li S, Mayhew SD, Kourtzi Z. Learning shapes the representation of behavioral choice in the human brain. Neuron. 2009; 62:441–52. [PubMed: 19447098]

Little DM, Thulborn KR. Correlations of cortical activation and behavior during the application of newly learned categories. Brain Res. Cogn. Brain Res. 2005; 25:33–47. [PubMed: 15936179]

Logothetis NK, Sheinberg DL. Visual object recognition. Annu. Rev. Neurosci. 1996; 19:577–621. [PubMed: 8833455]

Love BC, Medin DL, Gureckis TM. SUSTAIN: a network model of category learning. Psychol. Rev. 2004; 111:309–32. [PubMed: 15065912]

Lynd-Balta E, Haber SN. The organization of midbrain projections to the ventral striatum in the primate. Neuroscience. 1994; 59:609–23. [PubMed: 7516505]

Mahon BZ, Caramazza A. Concepts and categories: a cognitive neuropsychological perspective. Annu. Rev. Psychol. 2009; 60:27–51. [PubMed: 18767921]

Martin A. The representation of object concepts in the brain. Annu. Rev. Psychol. 2007; 58:25–45. [PubMed: 16968210]

McClelland, JL. How far can you go with Hebbian learning, and when does it lead you astray?. In: Munakata, Y.; Johnson, MH., editors. Processes of Change in Brain and Cognitive Development: Attention and Performance XXI. Oxford Univ. Press; Oxford: 2006. p. 33-69.

McClelland J, McNaughton B, O'Reilly R. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev. 1995; 102:419–57. [PubMed: 7624455]

Meeter M, Radics G, Myers CE, Gluck MA, Hopkins RO. Probabilistic categorization: How do normal participants and amnesic patients do it? Neurosci. Biobehav. Rev. 2008; 32:237–48. [PubMed: 18226722]

Merchant H, Zainos A, Hernández A, Salinas E, Romo R. Functional properties of primate putamen neurons during the categorization of tactile stimuli. J. Neurophysiol. 1997; 77:1132–54. [PubMed: 9084587]

Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. J. Neurophysiol. 2008; 100:1407–19. [PubMed: 18562555]

Middleton FA, Strick PL. Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. Science. 1994; 266:458–61. [PubMed: 7939688]

Middleton FA, Strick PL. Basal-ganglia `projections' to the prefrontal cortex of the primate. Cereb. Cortex. 2002; 12:926–35. [PubMed: 12183392]

Miller, EK.; Buschman, TJ. Rules through recursion: how interactions between the frontal cortex and basal ganglia may build abstract, complex rules from concrete, simple ones. In: SB; Wallis, JD., editors. The Neuroscience of Rule-Guided Behavior. Oxford Univ. Press; Oxford: 2007. p. 419-40.

Miller EK, Cohen JD. An integrative theory of prefrontal function. Annu. Rev. Neurosci. 2001; 24:167–202. [PubMed: 11283309]

Montague PR, Hyman SE, Cohen JD. Computational roles for dopamine in behavioural control. Nature. 2004; 431:760–67. [PubMed: 15483596]

Moody TD, Bookheimer SY, Vanek Z, Knowlton BJ. An implicit learning task activates medial temporal lobe in patients with Parkinson's disease. Behav. Neurosci. 2004; 118:438–42. [PubMed: 15113271]

Muhammad R, Wallis JD, Miller EK. A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. J. Cogn. Neurosci. 2006; 18:974–89. [PubMed: 16839304]

Murray EA, Bussey TJ, Wise SP. Role of prefrontal cortex in a network for arbitrary visuomotor mapping. Exp. Brain Res. 2000; 133:114–29. [PubMed: 10933216]

Myers CE, Shohamy D, Gluck MA, Grossman S, Kluger A, et al. Dissociating hippocampal versus basal ganglia contributions to learning and transfer. J. Cogn. Neurosci. 2003; 15:185–93. [PubMed: 12676056]

Newsome WT, Mikami A, Wurtz RH. Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion. J. Neurophysiol. 1986; 55:1340–51. [PubMed: 3734859]

Nomura EM, Maddox WT, Filoteo JV, Ing AD, Gitelman DR, et al. Neural correlates of rule-based and information-integration visual category learning. Cereb. Cortex. 2007; 17:37–43. [PubMed: 16436685]

O'Reilly, RC.; Munakata, Y. Computational Explorations in Cognitive Neuroscience: Understanding the Mind. MIT Press; Cambridge, MA: 2000.

Otani S, Blond O, Desce JM, Crépel F. Dopamine facilitates long-term depression of glutamatergic transmission in rat prefrontal cortex. Neuroscience. 1998; 85:669–76. [PubMed: 9639264]

Parthasarathy HB, Schall JD, Graybiel AM. Distributed but convergent ordering of corticostriatal projections—analysis of the frontal eye field and the supplementary eye field in the macaque monkey. J. Neurosci. 1992; 12:4468–88. [PubMed: 1279139]

Pasupathy A, Miller EK. Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature. 2005; 433:873–76. [PubMed: 15729344]

Poldrack RA, Clark J, Pare-Blagoev EJ, Shohamy D, Creso MJ, et al. Interactive memory systems in the human brain. Nature. 2001; 414:546–50. [PubMed: 11734855]

Poldrack RA, Foerde K. Category learning and the memory systems debate. Neurosci. Biobehav. Rev. 2008; 32:197–205. [PubMed: 17869339]

Poldrack RA, Prabhakaran V, Seger CA, Gabrieli JDE. Striatal activation during acquisition of a cognitive skill. Neuropsychology. 1999; 13:564–74. [PubMed: 10527065]

Poldrack RA, Rodriguez P. How do memory systems interact? Evidence from human classification learning. Neurobiol. Learn. Mem. 2004; 82:324–32. [PubMed: 15464413]

Reber PJ, Gitelman DR, Parrish TB, Mesulam MM. Dissociating explicit and implicit category knowledge with fMRI. J. Cogn. Neurosci. 2003; 15:574–83. [PubMed: 12803968]

Reber PJ, Squire LR. Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. Behav. Neurosci. 1999; 113:235–42. [PubMed: 10357448]

Reber PJ, Stark CE, Squire LR. Cortical areas supporting category learning identified using functional MRI. Proc. Natl. Acad. Sci. USA. 1998; 95:747–50. [PubMed: 9435264]

Redgrave P, Gurney K. The short-latency dopamine signal: a role in discovering novel actions? Nat. Rev. Neurosci. 2006; 7:967–75. [PubMed: 17115078]

Romo R, Salinas E. Touch and go: decision-making mechanisms in somatosensation. Annu. Rev. Neurosci. 2001; 24:107–37. [PubMed: 11283307]

Rorie AE, Newsome WT. A general mechanism for decision-making in the human brain? Trends Cogn Sci. 2005; 9(2):41–43. [PubMed: 15668095]

Schultz W, Apicella P, Ljungberg T. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J. Neurosci. 1993; 13:900–13. [PubMed: 8441015]

Schultz W, Apicella P, Scarnati E, Ljungberg T. Neuronal activity in monkey ventral striatum related to the expectation of reward. J. Neurosci. 1992; 12:4595–610. [PubMed: 1464759]

Seger CA. Implicit learning. Psychol. Bull. 1994; 115:163–96. [PubMed: 8165269]

Seger CA. How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. Neurosci. Biobehav. Rev. 2008; 32:265–78. [PubMed: 17919725]

Seger CA, Cincotta CM. The roles of the caudate nucleus in human classification learning. J. Neurosci. 2005; 25:2941–51. [PubMed: 15772354]

Seger CA, Cincotta CM. Dynamics of frontal, striatal, and hippocampal systems during rule learning. Cereb. Cortex. 2006; 16:1546–55. [PubMed: 16373455]

Seger CA, Peterson E, Lopez-Paniagua D, Cincotta CM, Anderson CM. Dissociating the contributions of independent corticostriatal systems to visual categorization learning through the use of reinforcement learning modeling and Granger causality modeling. NeuroImage. 2010; 50:644–56. [PubMed: 19969091]

Seymour B, Daw N, Dayan P, Singer T, Dolan R. Differential encoding of losses and gains in the human striatum. J. Neurosci. 2007; 27:4826–31. [PubMed: 17475790]

Shallice T. Specific impairments of planning. Philos. Trans. R. Soc. Lond. B Biol. Sci. 1982; 298:199–209. [PubMed: 6125971]

Shohamy D, Myers CE, Grossman S, Sage J, Gluck MA, Poldrack RA. Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. Brain. 2004; 127:851–59. [PubMed: 15013954]

Shohamy D, Myers CE, Kalanithi J, Gluck MA. Basal ganglia and dopamine contributions to probabilistic category learning. Neurosci. Biobehav. Rev. 2008; 32:219–36. [PubMed: 18061261]

Shohamy D, Wagner AD. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron. 2008; 60:378–89. [PubMed: 18957228]

Sigala N, Logothetis NK. Visual categorization shapes feature selectivity in the primate temporal cortex. Nature. 2002; 415:318–20. [PubMed: 11797008]

Smith EE. The case for implicit category learning. Cogn. Affect. Behav. Neurosci. 2008; 8:3–16. [PubMed: 18405041]

Smith EE, Grossman M. Multiple systems of category learning. Neurosci. Biobehav. Rev. 2008; 32:249–64. [PubMed: 17904637]

Tanaka K. Inferotemporal cortex and object vision. Annu. Rev. Neurosci. 1996; 19:109–39. [PubMed: 8833438]

Thierry AM, Blanc G, Sobel A, Stinus L, Glowinski J. Dopaminergic terminals in the rat cortex. Science. 1973; 182:499–501. [PubMed: 4744179]

Vallabha GK, McClelland JL, Pons F, Werker JF, Amano S. Unsupervised learning of vowel categories from infant-directed speech. Proc. Natl. Acad. Sci. USA. 2007; 104:13273–78. [PubMed: 17664424]

Vogels R. Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. Eur J. Neurosci. 1999; 11:1239–55. [PubMed: 10103119]

Vogels R, Sary G, Dupont P, Orban GA. Human brain regions involved in visual categorization. Neuroimage. 2002; 16:401–14. [PubMed: 12030825]

Wallis JD, Anderson KC, Miller EK. Single neurons in the prefrontal cortex encode abstract rules. Nature. 2001; 411:953–56. [PubMed: 11418860]

Wallis JD, Miller EK. From rule to response: neuronal processes in the premotor and prefrontal cortex. J. Neurophysiol. 2003; 90:1790–806. [PubMed: 12736235]

Williams ZM, Eskandar EN. Selective enhancement of associative learning by microstimulation of the anterior caudate. Nat. Neurosci. 2006; 9:562–68. [PubMed: 16501567]

Wingard JC, Packard MG. The amygdala and emotional modulation of competition between cognitive and habit memory. Behav. Brain Res. 2008; 193:126–31. [PubMed: 18565602]

Wyttenbach RA, May ML, Hoy RR. Categorical perception of sound frequency by crickets. Science. 1996; 273:1542–44. [PubMed: 8703214]

Yin HH, Knowlton BJ. The role of the basal ganglia in habit formation. Nat. Rev. Neurosci. 2006; 7:464–76. [PubMed: 16715055]

Zeithamova D, Maddox WT, Schnyer DM. Dissociable prototype learning systems: evidence from brain imaging and behavior. J. Neurosci. 2008; 28:13194–201. [PubMed: 19052210]

COMPUTATIONAL FACTORS IN CATEGORY LEARNING

Generalized knowledge versus memory for specific instances

The complementary memory systems framework notes that generalized knowledge (e.g., the overall concept of a chair) conflicts with specific memories (e.g., one's own office chair) (O'Reilly & Munakata, 2000). Categorization learning emphasizes the acquisition of generalized knowledge about the world but also requires some specific representations, for example in the situation of arbitrary categories or in representing exceptions to general rules.

Fast versus slow learning

Fast learning has obvious advantages: One can learn to acquire resources and avoid obstacles faster and better than competitors. But fast learning comes at a cost; it does not allow the benefits that come from generalizing over multiple experiences, so by necessity it tends to be specific and error prone. For example, consider conditioned taste aversion: a one-trial and often erroneous aversion for a particular food. Extending learning across multiple episodes allows organisms to pick up on the regularities of predictive relationships and leave behind spurious associations and coincidences. This allows category formation by allowing learning mechanisms to identify the commonalities across different category members. We suggest that the brain balances the advantages and disadvantages of fast versus slow learning by having fast plasticity mechanisms (large changes in synaptic weights) in subcortical structures train slower plasticity (small weight changes) in cortical networks.
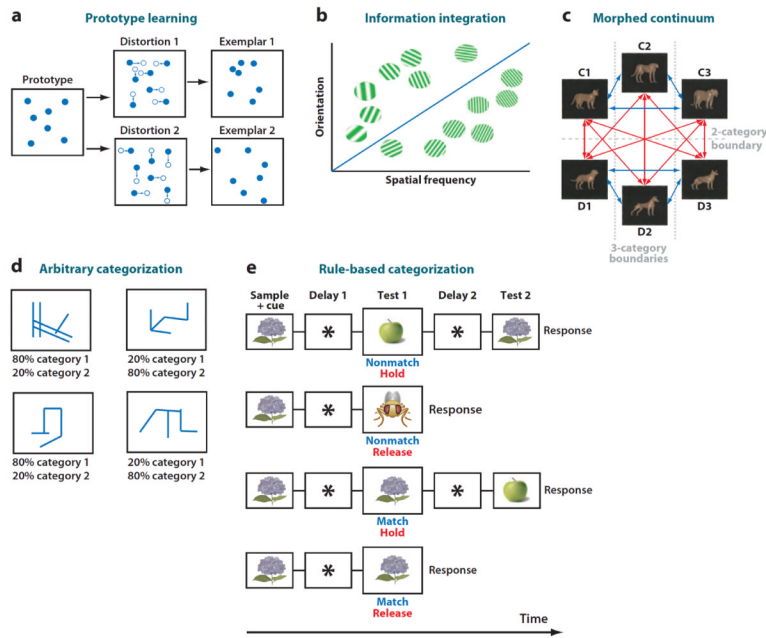
Figure 1.

Categorization tasks. (*a*) Dot pattern prototype learning. A prototypical stimulus is selected (*left*), and category exemplars (*right*) are formed by randomly moving dots. Large amounts of movement (*bottom*) result in high distortion stimuli; smaller amounts of movement (*top*) result in low distortion stimuli. (*b*) Information integration task. Stimuli are formed by varying two incommensurate features: angle from vertical and width of the bars. Illustrated is a diagonal decision bound between categories; to learn the categorization successfully, subjects must integrate the knowledge of angle and width. (*c*) Cat-dog categorization task. Stimuli are formed as continuous morphs along each of the lines between prototype stimuli. The categorical decision bound arbitrarily divides the continuous perceptual space into two or three domains, or categories. (*d*) Arbitrary categorization task. Each stimulus is individually probabilistically associated with the categories; stimuli within a category do not share identifying common features. (*e*) "Same - different" rule task. Monkeys responded on the basis of whether novel pairs of images matched or did not match, depending on which rule was in effect.
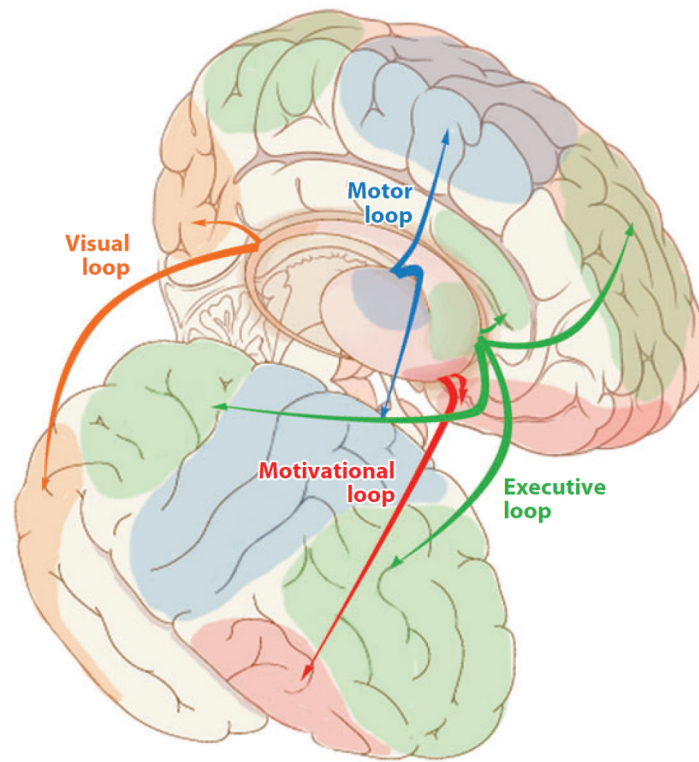
Figure 2.
Corticostriatal loops. The motor loop (*blue*) connecting the motor cortex with the posterior putamen. Executive loop (*green*) connects the prefrontal cortex and the parietal cortex with the anterior caudate nucleus. The motivational loop (*red*) connects the ventral striatum with the orbitofrontal cortex. The visual loop (*orange*) connects extrastriate and inferotemporal cortices with the posterior caudate nucleus.