Published in final edited form as:
Neuron. 2011 July 28; 71(2): 243–249. doi:10.1016/j.neuron.2011.05.040.

Differences between Neural Activity in Prefrontal Cortex and Striatum during Learning of Novel, Abstract Categories

Evan G. Antzoulatos and **Earl K. Miller***
The Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Summary

Learning to classify diverse experiences into meaningful groups, like categories, is fundamental to normal cognition. To understand its neural basis, we simultaneously recorded from multiple electrodes in the lateral prefrontal cortex and dorsal striatum, two interconnected brain structures critical for learning. Each day, monkeys learned to associate novel, abstract dot-based categories with a right vs. left saccade. Early on, when they could acquire specific stimulus-response associations, striatum activity was an earlier predictor of the corresponding saccade. However, as the number of exemplars was increasing, and monkeys had to learn to classify them, PFC began predicting the saccade associated with each category before the striatum. While monkeys were categorizing novel exemplars at a high rate, PFC activity was a strong predictor of their corresponding saccade early in the trial, before the striatal neurons. These results suggest that striatum plays a greater role in stimulus-response association and PFC in abstraction of categories.

Introduction

Virtually all animals have evolved some innate ability to group sensory inputs into useful categories like “food” and “mate”. Many animals can also learn new categories by abstracting diverse experiences. Humans are particularly adept at the latter; our brains seem predisposed to quickly learn the important commonalities among diverse items (e.g., “tool” or “pub”) which can then be used to recognize and interpret new experiences. As effortless as abstraction seems to be in neurotypical individuals, it can be compromised in neurological conditions. Take, for example, Temple Grandin, an individual with high-functioning autism, who has difficulty learning abstractions. She reports having no abstracted prototypes of, say, “dogs”, but, instead, retrieves from memory numerous individuals (Grandin, 2006).

There are many types of categories, from simple rule-based, to very complex and abstract. Several brain areas are involved, depending on the material to be categorized and the employed strategy (Ashby and Maddox, 2010; Seger and Miller, 2010). Human imaging studies have indicated activation of prefrontal cortex (PFC) and striatum (STR) in some types of category learning (Reber et al., 1998; Seger et al., 2000; Vogels et al., 2002). Although PFC plays a well-documented role in executive functions (Miller and Cohen, 2001), the role of STR in category learning is less intuitive: It is primarily known to be

*To whom correspondence should be addressed (ekmiller@mit.edu).

important for action selection and habit formation (Graybiel, 2005; Seger, 2008). A more detailed understanding of their roles in category learning may come from neuronal studies in monkeys. Several studies report that neurons in the monkey frontal and temporal cortex and STR show selectivity for learned stimulus groupings (Cromer et al., 2010b; Everling et al., 2006; Freedman et al., 2001; Kiani et al., 2007; Muhammad et al., 2006; Roy et al., 2010; Sigala and Logothetis, 2002; Sripati and Olson, 2009; Vogels, 1999). However, because category–related neural activity in monkeys has been examined only after extensive training, the respective roles of PFC and STR in the learning of new categories are not yet understood.

We designed a task in which monkeys could rapidly learn new abstract categories within a single experimental session, while we recorded from multiple electrodes simultaneously in lateral PFC and dorsal STR. It was based on a test of human category learning, the prototype distortion paradigm (Posner et al., 1967). It employs a large collection of constellations of dots by distorting the positions of a prototype pattern. Following experience with enough exemplars, humans learn (without seeing the prototypes) to abstract each category and categorize novel exemplars. This has been used in human (Posner et al., 1967), monkey (Smith et al., 2008), and pigeon (Blough, 1985) studies for the past 40 years but never with neuron recordings. Subjects can learn to distinguish between two categories (“A vs. B”) or one (“A vs. not A”). We used the A vs. B because amnesic patients display impaired performance in it, suggesting that it engages more “conscious” memory systems (Squire and Knowlton, 1995; Zaki et al., 2003).

Each training session began with a single exemplar per category. Monkeys learned them as specific stimulus-response (S-R) associations. We added more and more novel exemplars as learning progressed. This design (Katz and Wright, 2006), requires animals to learn the categories (or fail) because sooner or later, they would be confronted with too many novel exemplars (> 100; Fig 1C) to sustain above-chance performance via S-R learning. In our task, each category was always associated with a saccade direction. This was necessary for monkeys to learn new categories in a single experimental session. We used the development of saccade-related activity during training as an index of learning, as in prior studies (Asaad et al., 1998; Cromer et al., 2010a; Pasupathy and Miller, 2005). The prime interest was the early-trial activity, well before the animal's “go” signal. Changes in the early-trial neural activity presumably reflected the monkeys' improvement at classifying each exemplar into one of the categories, as expected with learning.

Results

Behavioral evidence for category abstraction

Every day, two monkeys were trained on a new pair of categories (Fig 1A). The exemplars of each category were created by shifting each of 7 dots at a random direction and distance from its prototypical location (Fig. 1B; Posner et al., 1967; Squire and Knowlton, 1995; Vogels et al., 2002). The distinction between the two categories was, therefore, not based on a simple rule. The monkeys' task was to learn to associate, by trial and error, each category with a saccade to a right vs. left target. The training session began with one exemplar per saccade direction, and once performance criterion was met (80% correct in last 20 trials) the two exemplars were supplemented with another two (Fig. 1C). Thus, at least during the first two blocks, behavior could be supported by learning specific S-R associations between individual exemplars and saccades. On Block 3, the two exemplars that were first introduced in Block 2 (which we term “familiar”) were supplemented with another 6 novel exemplars to double the total number from Block 2 (the original 2 exemplars from Block 1 were no longer shown, thus leading to a total number of 8 exemplars in Block 3). The same procedure was repeated on each subsequent block: Block n included the exemplars that were

novel in Block n-1 plus enough novel ones to bring the total number to 2^n (Fig 1C, and Suppl. Information). By Block 8, the last block in the sequence, monkeys were tested from a pool of 256 exemplars, 66% of which (i.e., 168) were novel.

We examined the average performance for the novel exemplars in each block across all days (Fig. 2A). Performance in Block 1 started from chance levels (50% correct), as expected, but showed a steep learning curve, consistent with S-R association learning. On every later block, behavioral performance on the novel exemplars tended to show a less steep learning curve, until it reached asymptote. In fact, by the 5^{th} block and beyond, monkeys' performance was high and stable even though they had to classify more and more novel exemplars. Indeed, the last few blocks largely consisted of novel exemplars, with the monkeys correctly classifying them on their first presentation: the hallmark of categorization. It is worth noting that category abstraction was not an inevitable consequence of experience. On a few sessions (5/24), monkeys failed to fully learn the categories and complete the task. They stayed at a low level of performance even though they remained motivated to try. In order to analyze the neurophysiological basis of category learning, we focused all our analyses on the sessions in which monkeys showed successful category learning and completed all 8 blocks (n=19).

We examined the extent to which the animal's saccade choice could be attributed to the individual exemplar vs. the category via an information theoretic approach (Fig. 2B; Shannon, 1948). We computed the shuffle-corrected mutual information between saccade choice and the exemplars tested in each block, as well as between saccade choice and the categories (see Suppl. Information). Mutual information between two variables (e.g., saccade choice and exemplar) quantifies the dependence between the 2 variables and reflects the fact that, if, say, left saccade is dependent on exemplar A, there is a higher probability to observe left saccade and exemplar A as a joint event than it is to observe each of these 2 events independently. The information that saccade choice carried about individual exemplars showed a transient rise in the first few blocks, but quickly decayed to a very low asymptote (approx. 0.08±0.01 bit [SEM]; Fig 2B). Indeed, on the last few blocks each exemplar was rarely repeated and thus information to be gained from its identity was diminished (Fig. S1). In contrast, information about the category, although it started from the same levels (0.135 ±0.058 bit) since category and exemplar were the same in the first 2 blocks, quickly rose to significantly higher levels than the exemplar information (Fig 2B; asymptoting at approx. 0.5 bit). A 2-way ANOVA (block number vs. variable) revealed significant interaction between block number and variable (i.e., exemplar vs. category; $p<2\times10^{-4}$). This means that as the number of exemplars was increasing, saccade choice became better predicted by category than the individual exemplars.

The number of different exemplars showed a progressive increase across blocks, and its average saturated after block 6 (at 23.53 ±2.41), indicating that the animals were reaching criterion even before all exemplars had been encountered in each block (Fig.2C left). Similar patterns across blocks were also observed in the probability of exemplar repetition and in the number of trials to criterion (i.e., both decreased across blocks; see Suppl. Information). We focused subsequent analyses on the novel exemplars of each block because we were interested in category learning per se and because familiar exemplars constituted only a small percentage of the trials, insufficient for reliable neurophysiological analysis (see Fig 2C right). Because of the variability in block length, we analyzed neural information across a 16-trial segment of novel exemplars from the start of each block.

The first two blocks involved learning of single specific exemplar-saccade associations. We pooled them as the “S-R Association” phase. During S-R Association, saccadic choice of novel exemplars on the first presentation was at chance (median of 50%, Inter-Quartile

Range: 50%). Category learning presumably took place from block 3 on, once the animals were exposed to multiple exemplars from each category. However, we also had to distinguish between *learning* and *performance* of the categories. To determine the first block in which performance relied on the newly learned categories, we set an operational criterion: a minimum of 75% success on the trials in which monkeys saw each novel exemplar for the very first time (for each category separately). The median block number that first met this criterion was 5. We pooled the first two blocks after criterion as the "Category Performance" phase. During Category Performance, a median of 94% (IQR: 13%) of novel exemplars were classified correctly on their first presentation. The pooled blocks between these phases (median number of 2 blocks) we classified as "Category Acquisition" phase. A median of 83% (IQR: 29%) of novel exemplars were classified correctly on their first presentation during Category Acquisition. This separation of experimental phases allowed us to collapse the block dimension and pool data from multiple blocks, as done previously (Cromer et al., 2010a; Pasupathy and Miller, 2005).

Neural activity during category learning

We report neurophysiological results from analyses of all simultaneously recorded neurons in the lateral PFC (344 neurons) and dorsal striatum (STR; 256 neurons; Fig. S2). Neural activity in STR was recorded from the head and body of the caudate nucleus, as previously (Muhammad et al., 2006; Pasupathy and Miller, 2005). To avoid biasing neuron selection, we pooled analyses across all randomly recorded, well-isolated neurons. This allowed us to simultaneously track learning-related changes in activity across the two neural populations under identical conditions. We estimated category/saccade information for every neuron, using the d′ sensitivity index (Dayan and Abbott, 2001), in a sliding 2-dimensional window (across trials and time) similar to previous studies (Cromer et al., 2010a; Pasupathy and Miller, 2005). The population averages were transformed into z scores based on the respective randomization distributions. Unless otherwise noted, all reported p-values are based on permutation tests.

Figure 3 shows different measures of the temporal dynamics of neural information about category/saccade direction as a function of time during the correctly performed trials of the novel exemplars. Figure 3A shows an overall picture of the dynamics of neural information and behavior while Figures 3B and 3C show more specific measures (i.e., average information and rise-time). In general, category-learning related (saccade-direction predicting) signals were stronger in PFC than STR. During S-R Association, STR predicted the behavioral response earlier in the trial than PFC (shortly after the exemplar onset; see below). During Category Acquisition though, early-trial category/saccade-predictive signals weakened in STR, while in PFC they strengthened and appeared earlier than in STR. During Category Performance, after the categories had been abstracted, early-trial signals in PFC remained stronger and earlier than in STR. To quantify the temporal dynamics of information, we measured the amount of saccade-direction information early vs. late in the trial. We also used rise-time (Pasupathy and Miller, 2005) to measure when saccade-direction information first reached considerable strength on each trial (half-maximum). Two-Way ANOVA (3 experimental phases × 2 neural populations) revealed significant interaction ($p<10^{-6}$) for each of these 3 measures (i.e., early-trial information, late-trial information, and rise-time). Details on the post-hoc comparisons are provided below. Single neuron examples and population averages are in Fig. S3.

During S-R association-based performance, striatum predicts the behavioral response before PFC

The bottom row of Figure 3A shows changes in neural information during the initial two blocks when there was a small number of exemplars and monkeys learned specific S-R

associations. While the PFC showed strong information about the saccade around its execution at the end of the trial, STR activity was a stronger predictor of the forthcoming saccade direction early in the trial (during and shortly after the exemplar). This is when monkeys, based on learning a few S-R associations, could first start to predict the saccade that would lead to reward. Rise-time in STR was an average of 130.7 ms (±12.9, SEM) across trials of the S-R Association phase. This is in contrast to PFC, where average rise-time was significantly later, at 822.1 ms (±128.2, $p<5\times10^{-4}$, Fig 3B). Likewise, during the early-trial epoch (exemplar display and the first half of the delay), information about the forthcoming saccade was significantly higher in STR (1.90 ±0.04) than PFC (1.0 ±0.04, $p<10^{-4}$, Fig 3C, left). In contrast, late in the trial (second half of the delay and during saccade execution), saccade information was stronger in PFC (2.44 ±0.05) than STR (0.83 ±0.05, $p<10^{-4}$, Fig 3C right). These results indicate that STR played a more leading role than PFC when performance relied on specific S-R associations.

A comparison of correct and error trials during the S-R phase is shown in Figure 4. In both cases, monkeys execute a right or left saccade. If activity reflects a motor signal per se, information should be equal on both. Yet, early-trial information in STR was greatly reduced on error vs correct trials (0.02 ±0.04, $p<10^{-4}$; Fig. 4A, B). It was lower when correct and error trials were pooled together and classified according to exemplar (1.38 ±0.04, $p<10^{-4}$; Fig. 4C), or saccade (0.70 ±0.03, $p<10^{-4}$; Fig. 4D). There was also a decrease of PFC saccade information, late in error trials (error trials alone: 0.85 ±0.04, $p<10^{-4}$; correct & error trials by exemplar: 0.70 ±0.05, $p<10^{-4}$; correct & error trials by saccade: 1.68 ±0.06, $p<10^{-4}$). The lower information on error trials indicate the STR and PFC are not reflecting a saccade motor plan per se (including “guesses”), but rather are involved in learning the correct saccade. The saccadic motor plan might have been generated and maintained elsewhere.

During category acquisition, PFC starts predicting the behavioral response earlier than STR

During the Category Acquisition phase, monkeys were confronted with increasingly larger numbers of novel exemplars (Fig. 1C) and had to move beyond simple S-R association and associate the right and left saccades with each category rather than individual exemplars. Performance was maintained at a high level and improved even though with each block an increasing proportion of novel exemplars was introduced (Fig 3A, middle row). During this phase, strong early-trial, saccade-predicting activity in PFC first appeared. This was reflected in the sharp reduction in rise-time (Fig 3B) and increase in saccade-direction information in the early-trial PFC activity, relative to S-R Association ($p<0.005$ for rise-time and $p<10^{-4}$ for information magnitude, Fig. 3C). In contrast, early saccade-predicting signals became weaker in STR, although still apparent, especially for the first half of Category Acquisition trials (Fig 3A). There was significant increase in STR rise-time ($p<0.005$; Fig 3B), and a sharp decrease in early-trial, saccade-direction information ($p<10^{-4}$; Fig 3C). The average rise-time in PFC (253.6 ms ±24.2) was significantly shorter than that in STR (476.4ms ±62.7, $p<0.01$), and early-trial information was significantly stronger in PFC (1.96 ±0.04) than STR (1.16 ±0.04, $p<10^{-4}$). Late in the trial, around saccade execution, saccade-related information was also significantly stronger in PFC (2.04 ±0.05) than STR (1.67 ±0.04, $p<10^{-4}$, Fig 3C).

After the monkeys reached the category learning criterion (Category Performance Phase), they were able to correctly categorize novel exemplars the first time they saw them. Early in trial, saccade-predicting information remained relatively strong in PFC (rise-time: 352.1 ms ±24.1), significantly earlier than in STR (729.3ms ±140.6, $p<0.01$; Fig. 3B). Early-trial category information in PFC (1.81 ±0.04) was also significantly stronger than in STR (1.34 ±0.04, $p<10^{-4}$; Fig. 3C). In contrast, saccade-related activity late in the trial, around saccade

execution, was similar in PFC (2.03 ±0.05) and STR (2.05 ±0.05, p=0.72). Within PFC, there was a small but significant decrease in early-trial information ($p<0.01$) and increase in rise-time ($p<0.05$), compared to the Category Acquisition phase. Within STR, in turn, there was no significant change in rise-time (p=0.12) but a significant increase in early-trial information ($p<0.005$), when compared to the Category Acquisition phase. These results suggest that, in contrast to the S-R phase of the session, PFC played a more leading role in learning and performing the categories than did STR, which only showed category/saccade information with longer latency.

Discussion

Monkeys learned to categorize novel exemplars from two new categories over a single experimental session by associating the exemplar category with a right vs. leftward saccade. We structured the animals' experience in such a way as to enforce a transition from an S-R association strategy to an abstract categorization strategy. Early in learning, when there were few exemplars, they could memorize specific S-R associations. Increasing the number of novel exemplars with learning encouraged them to abstract the “essence” of each category as the number of possible S-R associations became overwhelming. By the end of learning, monkeys were categorizing novel exemplars at a high level, even when seeing them for the very first time and having never seen the prototypes.

In the S-R Association phase, early-trial activity in STR more strongly predicted the behavioral response (saccade direction) for each exemplar than did PFC activity. Information in PFC was stronger than in STR late in the trial, around the time monkeys executed the corresponding response. However, robust changes were observed as soon as the animals were exposed to the diversity of the exemplars and started abstracting the categories: Early-trial saccade-predicting activity became stronger in PFC and weaker in STR. By the time the categories were learned, PFC activity predicted the correct behavioral response both stronger and earlier than STR, which instead showed increased information during the delay interval and late in the trial, around the time of motor planning/execution. Thus, with category learning, PFC signals shifted earlier in the trial (around the time monkeys could extract the exemplar's category and predict the behavioral response) whereas STR signals shifted later in the trial (around the time of saccade planning and execution). The apparent increase of category information in STR, along with the observed increase in rise-time and decrease of information in PFC, during the Category Performance phase may indicate that steady-state categorization was becoming habitual, as the animals were becoming more familiar with the categories.

We previously examined the same PFC and STR regions in monkeys performing non-category, pure S-R learning tasks (Asaad et al., 1998; Cromer et al., 2010a; Pasupathy and Miller, 2005). Like the current study, there was rapid development of learning-related signals in STR, but in contrast to the current study, they also developed in PFC, albeit lagging several trials behind those in STR (Pasupathy and Miller, 2005). In this study, we only saw learning-related, short-latency signals in the PFC after S-R association learning, during Category Acquisition, even though we previously found that during novel S-R learning, this activity can develop in PFC in as little as five correct trials (Cromer et al., 2010a). PFC activity does not simply reflect a correlation with the animal's level of performance per se. Our monkeys reached a high-level of performance during the S-R phase with little apparent early-trial saccade-predicting PFC activity; they also showed an improvement in behavior during the Category Performance phase when there was actually a small decrease in PFC information (perhaps due to increasing familiarity with the categories). The differences between studies, as well as the functional relationship between the PFC and STR, could be related to the dependence of PFC activity on task demands. The

monkeys had experience with each learning task and thus could have adopted different long-term strategies, depending on whether the task involved single S-R associations (Cromer et al., 2010a), learning and reversal of S-Rs (Pasupathy and Miller, 2005), or category learning (present study).

One clue to the PFC-STR functional relationship may lie in the anatomical loops connecting frontal cortex, striatum and basal ganglia. Our study targeted the PFC-dorsal striatum "associative" loop. We hypothesized that the faster plasticity in STR first acquires associations and then "trains" slower learning mechanisms in the PFC (Pasupathy and Miller, 2005). During learning of abstractions like categories, STR could first acquire specific associations. Category acquisition could occur as the output of the basal ganglia trains cortical networks which, by virtue of their slower plasticity, can pick up on the common features across specific exemplars and form abstract representations of the category (Miller and Buschman, 2008; Seger and Miller, 2010). This is consistent with observations that familiar abstract rules are represented more strongly and with a shorter latency in the frontal cortex than STR of monkeys (Muhammad et al., 2006) and thus were more likely stored in the PFC. Our finding that the strongest learning-related signals in STR appeared early in (S-R) learning, followed by stronger engagement by the PFC during and after category acquisition, is consistent with this hypothesis. In short, although our results do not preclude an important role for STR in the acquisition of abstractions by the PFC, they suggest greater engagement of PFC than STR neural mechanisms during category learning per se.

Experimental Procedures

Animals

Data were collected from two macaque monkeys, taken care of in accordance with the National Institutes of Health guidelines and the policies of the Massachusetts Institute of Technology Committee for Animal Care.

Task

Trials began when the animal maintained fixation on a central target for 0.7 s. Following fixation, a randomly chosen exemplar from either category was presented for 0.6 s (cue). Trials from both categories were randomly interleaved throughout the session. After the cue offset, there was a 1-s delay interval, followed by the saccade epoch, during which the fixation target was extinguished and two saccade targets appeared left and right of the center of fixation. The animal had to make a single, direct saccade to the correct target within 1 s for reward. Exemplars comprised of static constellations of 7 randomly located dots, generated as intermediate-level distortions of the corresponding prototype (see Suppl. Information).

Neurophysiology

Simultaneous recordings from PFC and STR were performed using two multi-electrode (8-16) arrays, which were lowered at different sites every day. Spikes were sorted offline, using principal component analysis. All computations were done on MATLAB. Neural information was computed using the d′ sensitivity index, i.e., the absolute difference in average firing rate between two conditions, normalized to their pooled standard deviation, and was calculated along a trial × time sliding window (10 trials × 100 ms). Unless otherwise noted, only correct trials were used for neurophysiological analyses. To correct for sampling bias, we randomly shuffled the trials between the two categories 1,000 times and calculated the population average information for the corresponding trial-time bin for each permutation. The observed population average was subsequently transformed into a z

score, based on the 1,001 (incl. the observed one) permutations. For permutation tests, we randomly shuffled the data between two conditions (i.e., experimental phases or neural populations) 10,000 times, and quantified the probability of observing the given difference by chance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank S. Brincat, T. Buschman, J. Cromer, C. Diogo, D. Fioravante, V. Puig, J. Rose, J. Roy, M. Siegel, and M. Wicherski for helpful discussions and comments on the manuscript. They also thank K. MacCully and D. Ouellette for technical assistance, and J. Liu and M. Machon for their help in animal training.

This work was funded by the National Institutes of Mental Health (2R01MH065252-06), the Simons Foundation, and Richard and Linda Hardy.

References

Asaad WF, Rainer G, Miller EK. Neural activity in the primate prefrontal cortex during associative learning. Neuron. 1998; 21:1399–1407. [PubMed: 9883732]

Ashby FG, Maddox WT. Human category learning 2.0. Ann N Y Acad Sci. 2010

Blough DS. Discrimination of letters and random dot patterns by pigeons and humans. Journal of experimental psychology. Animal behavior processes. 1985; 11:261–280. [PubMed: 4009122]

Cromer JA, Machon M, Miller EK. Rapid Association Learning in the Primate Prefrontal Cortex in the Absence of Behavioral Reversals. J Cogn Neurosci. 2010a

Cromer JA, Roy JE, Miller EK. Representation of multiple, independent categories in the primate prefrontal cortex. Neuron. 2010b; 66:796–807. [PubMed: 20547135]

Dayan, P.; Abbott, LF. Theoretical Neuroscience. Cambridge: The MIT Press; 2001.

Everling S, Tinsley CJ, Gaffan D, Duncan J. Selective representation of task-relevant objects and locations in the monkey prefrontal cortex. Eur J Neurosci. 2006; 23:2197–2214. [PubMed: 16630066]

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representation of visual stimuli in the primate prefrontal cortex. Science. 2001; 291:312–316. [PubMed: 11209083]

Grandin, T. Thinking in Pictures: My Life with Autism. 2nd. New York: Vintage Books; 2006.

Graybiel AM. The basal ganglia: learning new tricks and loving it. Curr Opin Neurobiol. 2005; 15:638–644. [PubMed: 16271465]

Katz JS, Wright AA. Same/different abstract-concept learning by pigeons. J Exp Psychol Anim Behav Process. 2006; 32:80–86. [PubMed: 16435967]

Kiani R, Esteky H, Mirpour K, Tanaka K. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J Neurophysiol. 2007; 97:4296–4309. [PubMed: 17428910]

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: Bunge, SA.; Wallis, JD., editors. Neuroscience of Rule-Guided Behavior. New York: Oxford University Press; 2008.

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

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; 18:974–989. [PubMed: 16839304]

Pasupathy A, Miller EK. Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature. 2005; 433:873–876. [PubMed: 15729344]

Posner MI, Goldsmith R, Welton KE Jr. Perceived distance and the classification of distorted patterns. J Exp Psychol. 1967; 73:28–38. [PubMed: 6047706]

Reber PJ, Stark CE, Squire LR. Cortical areas supporting category learning identified using functional MRI. Proc Natl Acad Sci U S A. 1998; 95:747–750. [PubMed: 9435264]

Roy JE, Riesenhuber M, Poggio T, Miller EK. Prefrontal cortex activity during flexible categorization. J Neurosci. 2010; 30:8519–8528. [PubMed: 20573899]

Seger CA. How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. Neurosci Biobehav Rev. 2008; 32:265–278. [PubMed: 17919725]

Seger CA, Miller EK. Category learning in the brain. Annu Rev Neurosci. 2010; 33:203–219. [PubMed: 20572771]

Seger CA, Poldrack RA, Prabhakaran V, Zhao M, Glover GH, Gabrieli JD. Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. Neuropsychologia. 2000; 38:1316–1324. [PubMed: 10865107]

Shannon CE. A Mathematical Theory of Communication. Bell System Technical Journal. 1948; 27:623–656.

Sigala N, Logothetis NK. Visual categorization shapes feature selectivity in the primate temporal cortex. Nature. 2002; 415:318–320. [PubMed: 11797008]

Smith JD, Redford JS, Haas SM. Prototype abstraction by monkeys (Macaca mulatta). Journal of experimental psychology. General. 2008; 137:390–401.

Squire LR, Knowlton BJ. Learning about categories in the absence of memory. Proc Natl Acad Sci U S A. 1995; 92:12470–12474. [PubMed: 8618923]

Sripati AP, Olson CR. Representing the forest before the trees: a global advantage effect in monkey inferotemporal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2009; 29:7788–7796. [PubMed: 19535590]

Vogels R. Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. Eur J Neurosci. 1999; 11:1239–1255. [PubMed: 10103119]

Vogels R, Sary G, Dupont P, Orban GA. Human brain regions involved in visual categorization. Neuroimage. 2002; 16:401–414. [PubMed: 12030825]

Zaki SR, Nosofsky RM, Jessup NM, Unverzagt FW. Categorization and recognition performance of a memory-impaired group: evidence for single-system models. Journal of the International Neuropsychological Society : JINS. 2003; 9:394–406. [PubMed: 12666764]

Highlights

- Gradually increasing category exemplars leads to abstraction of categories
- Striatum processes single exemplar-response associations
- Prefrontal cortex abstracts categories from diverse exemplars
- Both prefrontal cortex and striatum are involved in steady-state categorization

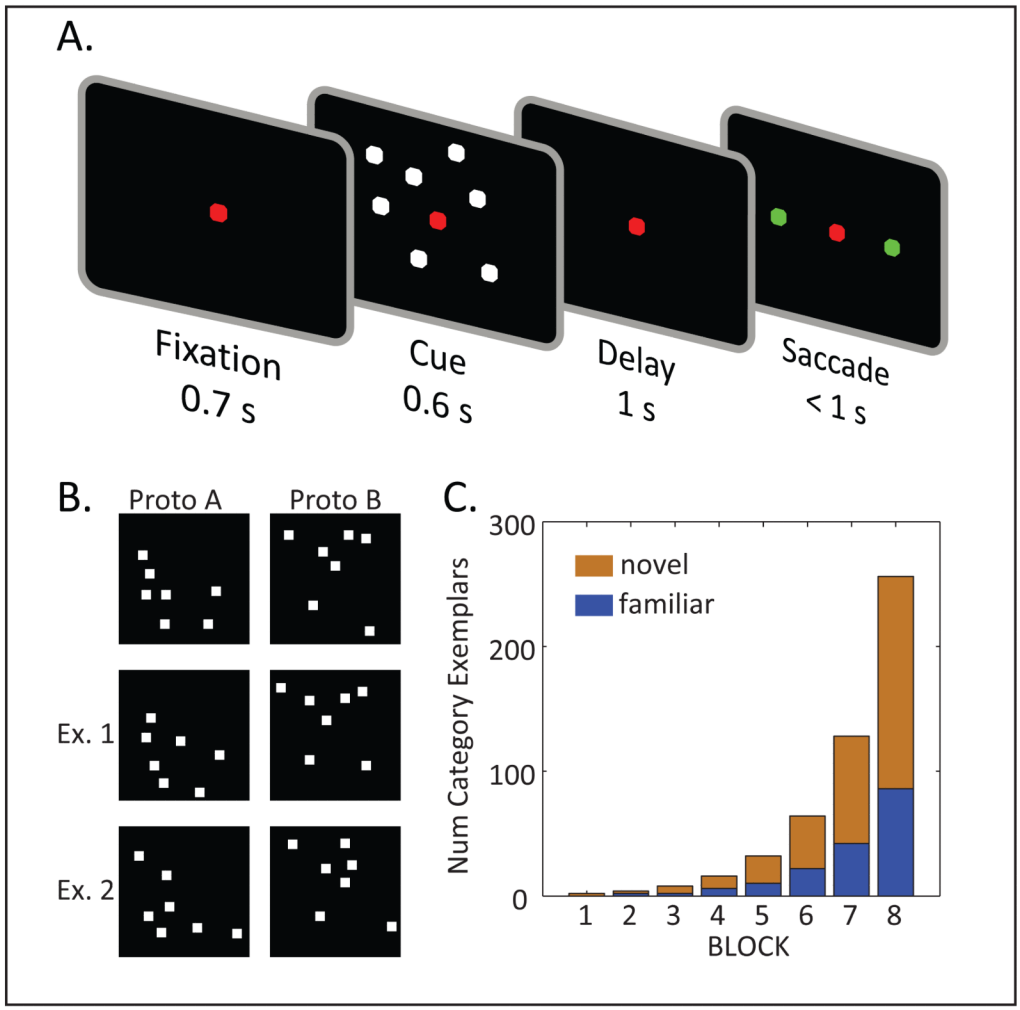

Figure 1. A Task of Abstract Category Learning

A. After an initial fixation period, a randomly chosen exemplar of category A or B was shown. Following a brief delay interval, the animal had to classify the exemplar by choosing between a saccade to the left or right target. B. Example stimuli: Top row of panels illustrates two example prototypes, and the other two rows illustrate two exemplars from each category. C. The first block included a single exemplar per category, and on every block, the number of category exemplars was doubled. All exemplars were included in the pool of only two consecutive blocks. “Familiar” (blue) indicates exemplars that were shared between each block and its previous one; “novel” (red) indicates those first introduced in that block.

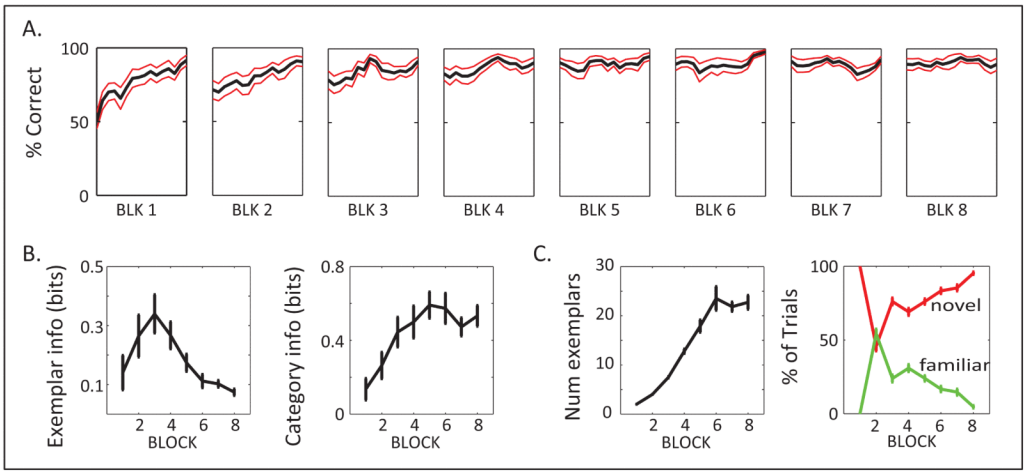

Figure 2. Behavioral Indices of Category Abstraction

A. Across-trial performance on novel exemplars is averaged across all sessions (n=19) for each block separately (First 16 trials per block; red lines indicate SEM). B. Average mutual information (bits) across blocks, between saccade choice and either exemplar identity (left), or category membership (right). C. Left: The average number of exemplars performed in each block gradually increased, until it reached asymptote in the last 3 blocks when the animals were reaching criterion before all exemplars could be tested. Right: Percentage of trials that tested novel exemplars (red line) vs. familiar exemplars (green line). Except for block 2, where both were at approx. 50%, the novel outnumbered the familiar exemplars. All error bars are SEM. (See also Fig. S1.)

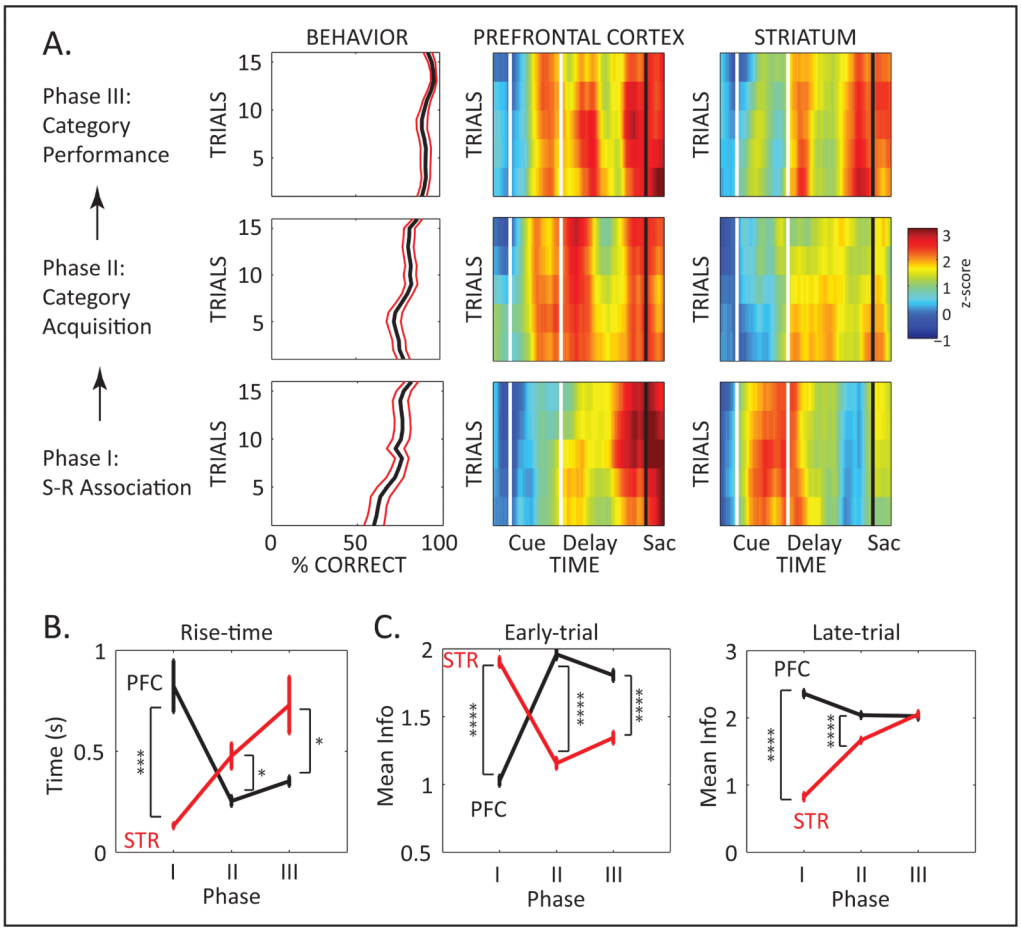

Figure 3. Dynamics of Information Processing in Prefrontal Cortex and Striatum during Category Abstraction

A. Left column of panels illustrates average behavioral performance (±SEM) across trials. The other 2 columns illustrate neural information for the PFC (middle) and striatum (STR; right) neural populations, across trials (y axis) and time (x axis). The fixation, cue, delay, and saccade epochs (also seen in Fig. 1A) are delimited by vertical lines. Information was computed in the same trial segment as behavioral performance, but in a sliding trial × time window. Bottom row of panels: S-R Association phase. Middle row: Category Acquisition phase. Top row: Category Performance phase. B. Average (±SEM) rise-time across trials, for PFC (black) and STR (red) in each of the three experimental phases shown in A. C. Average (±SEM) information in the PFC and STR neural populations in the early (left) and the late (right) epochs of the trial. (See also Figure S3.)

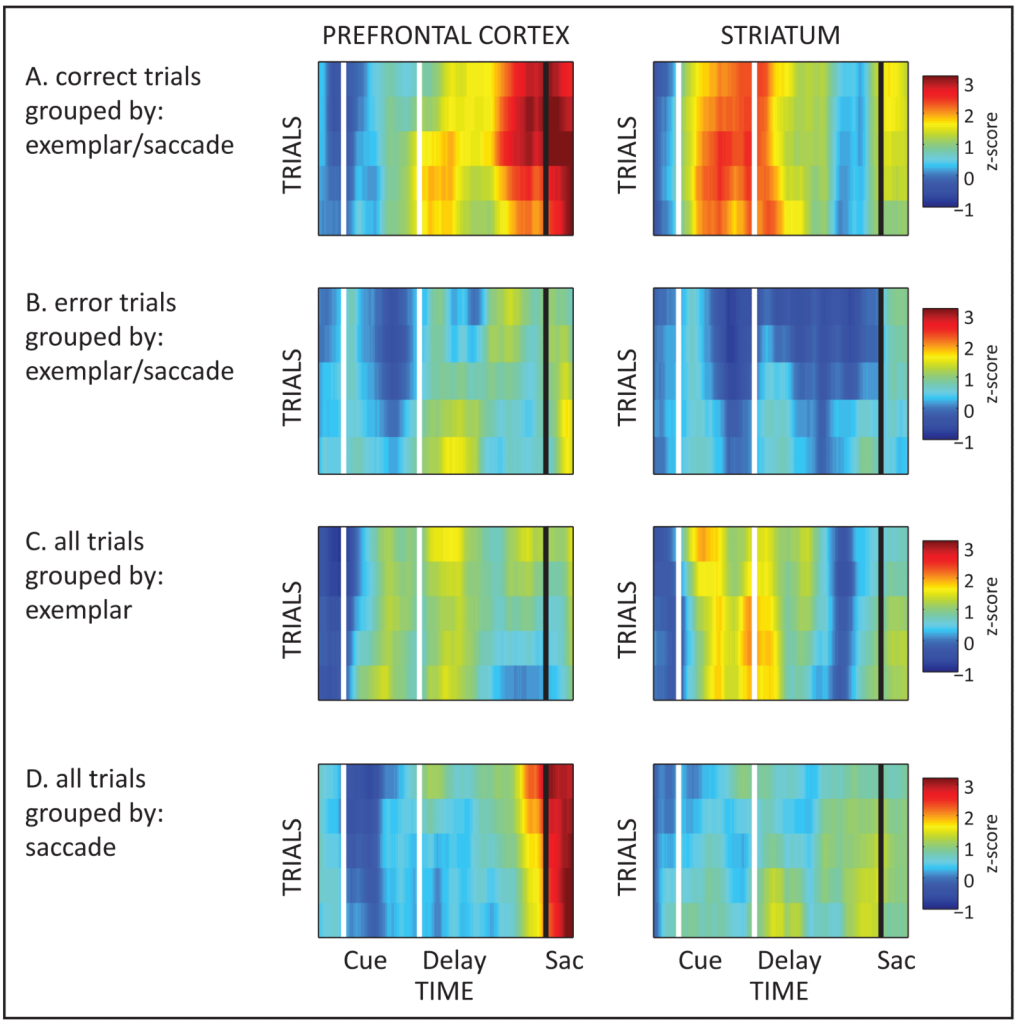

Figure 4. Error-trial analyses of S-R Association Phase

A. Same as bottom row of Fig. 3A: Neural information across trials and time in PFC (left) and STR (right) on correct trials only and error trials only (B). On both corrects and errors, monkeys execute a right or left saccade; the only difference are the exemplars. C. Same analysis, but on pooled correct and error trials. The trials are grouped according to the tested exemplar. D. Same as in C, but grouped according to saccade choice.