Behavioral/Systems/Cognitive

Prefrontal Cortex Activity during Flexible Categorization

Jefferson E. Roy,[1,4] Maximilian Riesenhuber,[5] Tomaso Poggio,[2,3,4] and Earl K. Miller[1,4]
[1]The Picower Institute for Learning and Memory, [2]Center for Biological and Computational Learning, and [3]McGovern Institute for Brain Research, [4]Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and [5]Department of Neuroscience, Georgetown University Medical Center, Washington, DC 20007

Items are categorized differently depending on the behavioral context. For instance, a lion can be categorized as an African animal or a type of cat. We recorded lateral prefrontal cortex (PFC) neural activity while monkeys switched between categorizing the same image set along two different category schemes with orthogonal boundaries. We found that each category scheme was largely represented by independent PFC neuronal populations and that activity reflecting a category distinction was weaker, but not absent, when that category was irrelevant. We suggest that the PFC represents competing category representations independently to reduce interference between them.

Introduction

Perceptual categorization is the ability to detect and store the relevant commonalties across items while ignoring their irrelevant differences. This is critical to normal thought because it allows us to recognize new items or old items in a new light. Impairment in visual category learning is found in a range of cognitive disorders such as autism (Hill, 2004) and schizophrenia (Kéri et al., 1999; Tan et al., 2006; Weickert et al., 2009).

Neural correlates of visual categories have been found in the temporal, parietal, and frontal cortices (Vogels, 1999; Freedman et al., 2001, 2002, 2003; Freedman and Assad, 2006; DeGutis and D'Esposito, 2007; Diester and Nieder, 2007). Their neurons often show the hallmarks of perceptual categorization: greater differences in activity in response to stimuli from different categories than in response to stimuli from the same category, regardless of their exact physical appearance. So far this has been demonstrated by training animals on static, fixed categories (e.g., a given stimulus was always a "dog" or a "cat"). This was a necessary first step in establishing the basic neural phenomenon and its distribution in higher-level cortex. However, one of the remarkable features of primate cognition is our great flexibility. We can classify the same object into different categories depending on the context of our current goal. For example, airplanes can be things that fly (like birds) or a mode of transportation (like trains). The nature of and substrate for this dynamic flexibility are not yet fully understood.

We investigated flexible categorization in the prefrontal cortex (PFC). The PFC is known to have neural correlates of shape-based categorization (Freedman et al., 2001, 2002, 2003; DeGutis and D'Esposito, 2007; Diester and Nieder, 2007) and is critical for cognitive flexibility (Goldman-Rakic, 1987; Fuster, 2000; Miller, 2000; Miller and Cohen, 2001; Poldrack and Rodriguez, 2004). We used the computer-generated cat-and-dog morph stimuli used in our previous studies (Freedman et al., 2001, 2002, 2003). A three-dimensional morphing system (Shelton, 2000) produced parametric blends (morphs) of four prototypes (two cat and two dog prototypes) (see Fig. 1*A*). The monkeys learned to categorize this image set using two orthogonal category schemes. One scheme divided the image set into cat-like and dog-like morphs (see Fig. 1*A*, left panel) and the other scheme grouped together different pairs of cat and dog prototypes (Fig. 1*A*, right panel). Our monkeys flexibly switched between these category schemes while we recorded neural activity from multiple electrodes in the PFC, targeting the same ventrolateral region as our prior work (Freedman et al., 2001, 2002, 2003).

Materials and Methods

Two *Macaca mulatta* (8–10 kg) were handled in accordance with National Institutes of Health guidelines and the Massachusetts Institute of Technology Committee on Animal Care. Eye movements were monitored using an infrared eye tracking system (Iscan) at a sampling rate of 240 Hz.

Stimuli. A large number of morph images were generated by varying the composition percentage of the two cat and two dog prototype images (Fig. 1*A*), using the vector differences between corresponding points (for more information, see Shelton, 2000). Morphs were linear combinations of these vectors added to the prototype. Stimuli from different categories differed along multiple features and were smoothly morphed (i.e., without sudden appearance of any feature). The stimulus space was divided into two different category schemes where the boundary lines were orthogonal. Each category scheme (Fig. 1*A*, scheme A in left panel and scheme B in right panel) designated two categories. An image was considered member of a category if it contained more than a 50% contribution from a prototype in that category. During training of both category schemes, the image set consisted of thousands of images generated from combinations of the four prototypes. For the recording sessions, we generated 28 images from six levels of combinations of each pair of prototypes (100:0, 80:20, 60:40, 40:60, 20:80, and 0:100) to be used as the sample images. The images of the six morph lines that span between the

Received Sept. 29, 2009; revised Feb. 25, 2010; accepted March 4, 2010.
This work was supported by National Institute of Mental Health Grant 2R01MH065252-06. We thank Timothy Buschman, Jason Cromer, David Freedman, Markus Siegel, and Marlene Wicherski for valuable comments, help, and discussions.
Correspondence should be addressed to Earl K. Miller, The Picower Institute for Learning and Memory, Department of Brain and Cognitive Science, 77 Massachusetts Avenue, Massachusetts Institute of Technology, Cambridge, MA 02139. E-mail: ekmiller@mit.edu.
DOI:10.1523/JNEUROSCI.4837-09.2010
 0270-6474/10/308519-10$15.00/0

prototypes are shown in Figure 1*A*. All images were 4.2 degrees in diameter and had identical color, shading, orientation, and scale.

Task. The monkeys performed a delayed match-to-category task (Fig. 1*B*). Monkeys initiated the trial by holding a lever and acquiring and maintaining fixation for 1000 ms. The color of the dot in the first 500 ms instructed which category scheme (blue for scheme A or red for scheme B) should be followed on that trial. For the last 500 ms, the dot was white in color. A sample image (chosen from the 28 described above) was presented for 600 ms, followed by a 1000 ms memory delay. Next, a test image was presented. If it matched the category of the sample image, monkeys released a lever to receive a juice reward. If it was a category nonmatch, they were to continue holding the lever through a second delay, which was followed by a match image requiring a response. To ensure that monkeys did not simply memorize specific stimulus–stimulus–response patterns, each day the test images for each category were randomly chosen from a pool of 150 morphs that were at least 70% of a prototype. Category scheme A/B and match/nonmatch trials were randomly interleaved and occurred at similar frequency. Monkeys maintained fixation throughout within a ±2 degree window that centered the stimuli on the fovea.

Both animals were first trained on category scheme A until their performance was 80% or better. This took ~6 months. Then the animals trained on category scheme B exclusively until they attained the same performance criteria; this took ~4 months. Only then were the two category schemes presented within the same session. At first, blocks of 20 trials of each scheme were used. Over the subsequent ~4 weeks, the number of trials in each block was reduced to one, at which point schemes could be randomly chosen.

Recording. Either 8 or 16 acute epoxy-coated tungsten electrodes (FHC) per recording session were lowered into the brain. Custom-made screw-driven microdrives were used to lower the electrodes, with each drive controlling two electrodes, through a plastic grid with 1 mm spacing (Freedman et al., 2001, 2002, 2003). Recording chambers were stereotaxically placed using MRI images and anatomical atlas (Paxinos et al., 2000) over the PFC (principal sulcus and anterior arcuate sulcus or areas 45, 46, and 12). Isolated neurons were not prescreened for task-related activity such as stimulus or category sensitivity. Rather, we recorded activity from every well isolated neuron we encountered. We were able to isolate an average of one or two neurons per electrode. The activity of 536 lateral prefrontal cortex neurons was recorded (333 from monkey O in 38 sessions and 203 from monkey L in 40 sessions). Reconstructions of the recording location are shown in Figure 1*C* for both monkeys. Note that the MRI images for monkey L were only available as printouts, which was sufficient for well placement, but we were unable to generate a three-dimensional reconstruction. We instead adjusted the reconstruction of monkey O to fit the measurements of monkey L. Waveforms were digitized and then stored for off-line sorting. Principal components analysis was used to sort the waveforms into individual neurons (Plexon).

Data analysis. Neuronal activity was averaged over four time epochs: baseline (400 ms before the sample image presentation), sample presentation (100–600 ms after sample onset), memory delay (300–1100 ms after sample offset), and test image presentation [100 ms after test onset to 2 SDs before each monkey's daily average reaction time (RT)]. The test image presentation epoch was chosen to avoid any influence of the behavioral response (monkey O: mean RT = 284 ms, mean test interval = 173 ms; monkey L: mean RT = 350 ms, mean test interval = 232 ms).

Neural activity was normalized first by subtracting the minimum activity during that epoch from the measured activity and then dividing by the difference of the maximum and minimum firing rates. This method was chosen to maximize the dynamic range of each neuron in each time interval and to be consistent with previous studies (Freedman et al., 2001, 2002, 2003). The findings were not affected if activity was not normalized or normalized to baseline activity. The average firing rate traces have been smoothed with a Savitzky–Golay filter (also called least-squares smoothing filters) with a weighting vector of 51 ms. We used standard statistical methods such as *t* tests and ANOVAs.

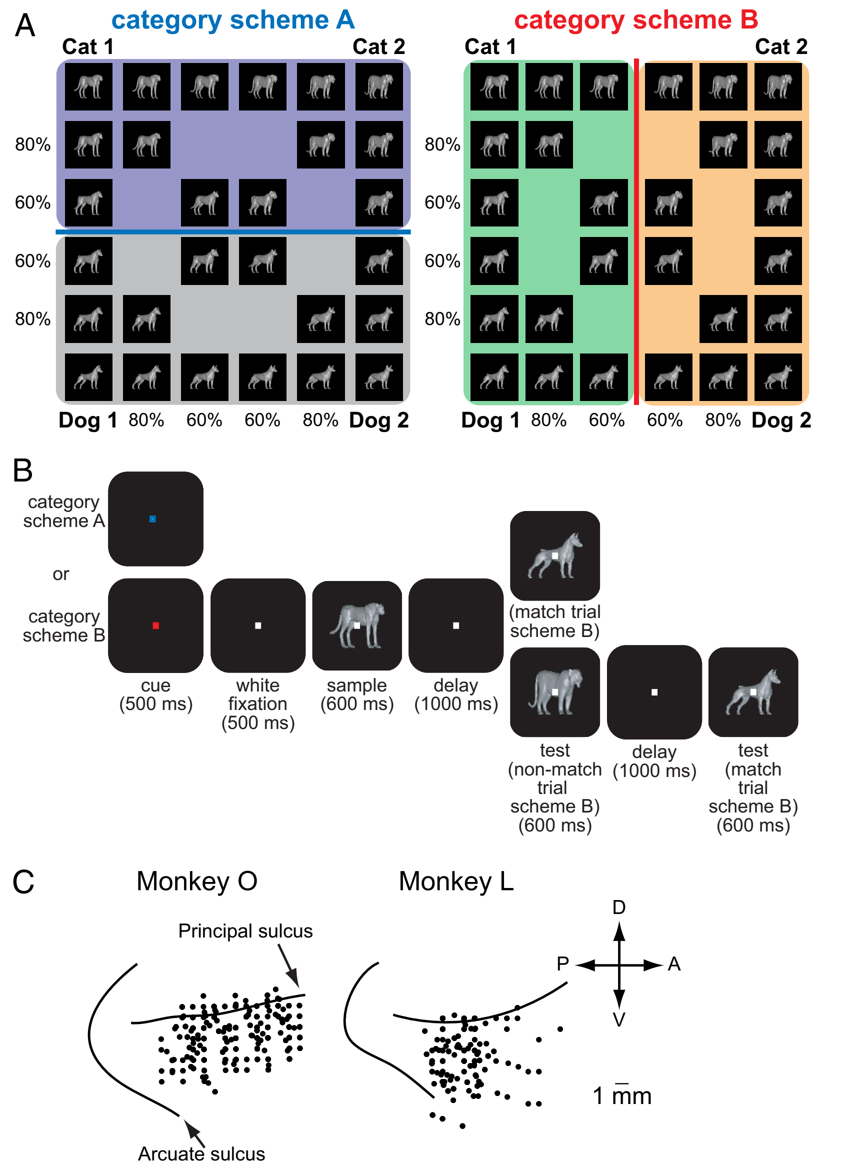

Figure 1. Stimuli and behavioral task. ***A***, An image stimulus set was generated by blending prototypes along six morph lines. Monkeys were taught to group the same images under two different category schemes. ***B***, Schematic diagram of the delayed match-to-category task. Each trial began with the monkeys fixating on a dot and holding a lever. The dot was briefly a color that cued the monkey as to which category scheme was relevant. The dot switched to white and a sample image appeared for 600 ms. Following a brief 1 s delay, a test image was presented. If the test image and sample image were of the same category, monkeys released a lever. Otherwise, monkeys continued to hold the lever through a delay until a matching image was displayed. ***C***, Anatomical locations of recording sites of category selective neurons in both monkeys. A, Anterior; P, posterior; D, dorsal; V, ventral.

A category index was generated to assess the strength of the category selectivity. We calculated each neuron's difference in average activity in response to pairs of images along the morph lines that crossed the category boundary (Freedman et al., 2001, 2002). The within-category difference (WCD) was defined by computing the absolute difference between the 100% and 80% morphs and the 80% and 60% morphs for both categories and averaging these values. The between-category difference (BCD) was computed by averaging the across-boundary differences between the 60% of one category and 60% of the other category. The distance between the images was identical at 20% for both the WCD and BCD. The index was calculated by dividing the difference between BCDs and WCDs by their sum and could range between -1 and 1. The more positive the index, the larger the difference in responses to images that are between categories as compared with within categories. The category index was calculated for the sample presentation and memory delay intervals independently.

To capture category effects in the activity of the population of neurons, we calculated correlation coefficients. Arranging the correlation coefficients into a matrix can reveal patterns of stimulus and category selectivity (Hegdé and Van Essen, 2006; Freedman and Miller, 2008). For each image, we used average activity of each neuron during both the sample and memory delay intervals (considered independently) and calculated correlations between each neuron's activity in response to a given image and every possible pairing of each other image. The coefficients reflect the degree of similarity of activity in response to the different images. If the activity level was similar to the two images, the correlation would be higher than if activity was different. Permutation tests (Manly, 1997) were used to assess whether average coefficient values were greater than those observed by chance (null hypothesis) or whether differences between values were significant. The permutation tests (repeated 5000 times) provided a null distribution with which the calculated values could be compared. For each permutation, the category assignments of the images were randomized and the correlation coefficients and subsequent average value were calculated. Significance (p value) was calculated by taking the number of calculated average values that were greater than the observed values and dividing by the total number of permutations.

The correlation coefficient data were used as the input for the multidimensional scaling (MDS) analysis. MDS plots the data, in this case each image, so that the distance between data points represents the similarity of the population of neural responses to the images. MDS was calculated using a nonlinear dimensionality reduction method (Tenenbaum et al., 2000) that aims to preserve any intrinsic geometry in the data (neighborhood function, $K = 14$). The average distances between images within each category and between categories were calculated.

Results

Behavior

Both monkeys were able to flexibly categorize the images at a high level of proficiency. Monkey O (Fig. 2*A*) and monkey L (Fig. 2*B*) were equally good (>80% correct) at categorizing images under both category schemes A and B. Monkey L's overall performance (mean $\pm$ SD) was 87 $\pm$ 19% correct which was slightly, but significantly ($p < 0.01$), worse than monkey O who performed at 89 $\pm$ 17% correct. Both monkeys correctly categorized images at >80% correct even when the images were close to the boundary lines (i.e., the 60% morphs). In their combined performance (Fig. 2*C*), there was no significant differences in performance between the two category schemes at each morph level (*t* test, $p > 0.01$).

General neural properties

The main task epochs of interest were the sample presentation and memory delay intervals. That was when the monkeys had to categorize the sample image and hold that information in short-term memory. The majority of neurons were responsive in that they showed a significant change in neural activity relative to their baseline firing rate during either the sample or memory delay epochs (425 of 536 or 79.3%; paired *t* test, $p < 0.05$, Bonferroni corrected).

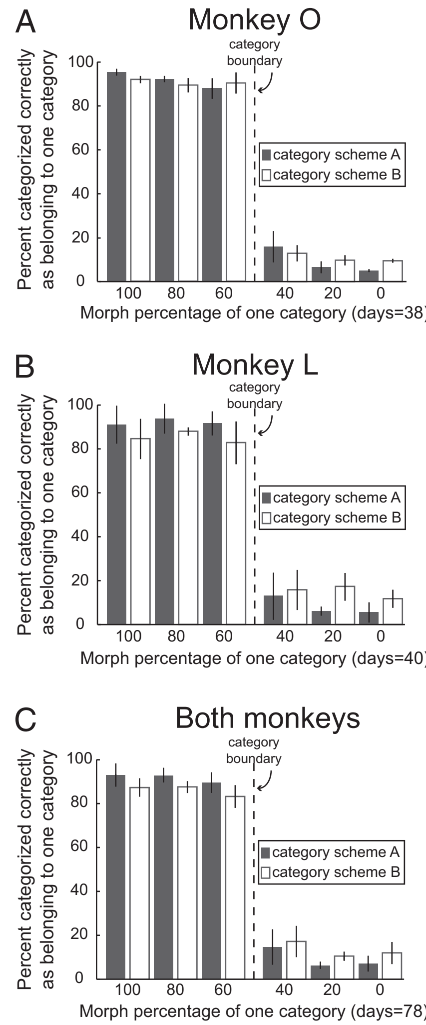

Figure 2. Behavioral performance of both monkeys. ***A***, ***B***, Monkeys O (***A***) and L (***B***) were proficient at categorizing images in both schemes (scheme A, filled bars; scheme B, open bars). ***C***, The performance of the two monkeys combined.

Our first step was to identify a population of neurons for further analysis by determining which neurons could potentially show category effects. To include the greatest number of potential category neurons, we compared, for each category scheme, the average neural activity in response to the 14 images of one category to the 14 images of the other category for that scheme (Fig. 1*A*). Many neurons (~38.4% or 206 of 536, 66 during the sample presentation, 92 during the memory delay, and 48 during both intervals) showed a significant difference in overall activity between the categories (*t* test, $p < 0.05$, Bonferroni corrected). We will refer to these neurons as "category sensitive," not category selective, because this population also could include neurons that show strong stimulus selectivity for one or more exemplars from one category. Analyses presented in the following sections will show that this population of neurons did show hallmarks of category effects (i.e., greater selectivity across than within categories and a sharp change in activity across category boundaries).

A key question was how neural information about the two different category schemes was distributed among neurons. There are two possibilities: at one end of the spectrum, neurons

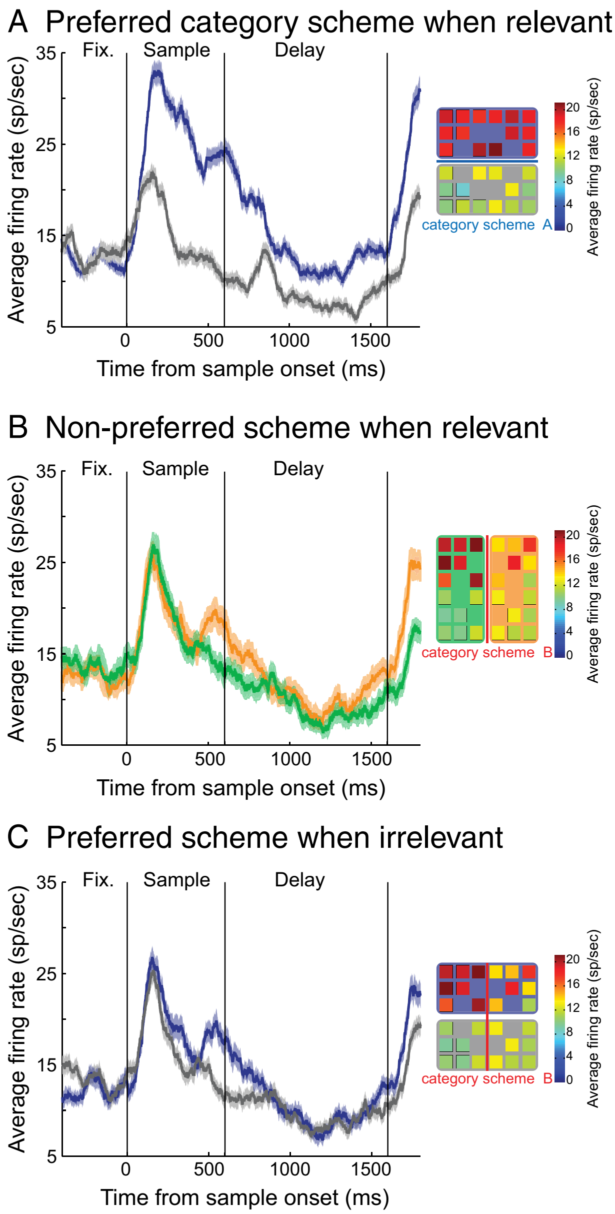

Figure 3. Category sensitivity of an example PFC neuron. ***A***, The neuron's average activity (mean ± SEM) in response to all images from the two categories under category scheme A. Insets indicate average responses to each image across the sample presentation and memory delay intervals. ***B***, The same neuron's activity when scheme B was relevant and data were sorted by that scheme. ***C***, The neuron carried less information about category scheme A when the animal was performing category scheme B. sp; Spikes; Fix., fixation epoch.

could be multitasking. Individual neurons could show selectivity for each category distinction under both schemes. 2) At the other end of the spectrum is specialization, completely separate neuron populations could show selectivity under the two different category schemes.

Category activity under different category schemes

We found that two categorical distinctions under the two category schemes were reflected by largely independent PFC neuron populations. That is, most neurons showed category selectivity under one category scheme, but not the other. An example neuron is shown in Figure 3*A*. It showed a clear difference in activity when the monkey was grouping the images under one category scheme (Fig. 3*A*). We will refer to this as the neuron's "preferred"

Table 1. Number of PFC neurons sensitive to each category scheme and trial interval

	Category scheme A	Category scheme B	Both schemes
Sample presentation	32 of 536 (6.0%)	21 of 536 (3.9%)	13 of 536 (2.4%)
Memory delay	36 of 536 (6.7%)	37 of 536 (6.9%)	19 of 536 (3.5%)
Both intervals	22 of 536 (4.1%)	9 of 536 (1.7%)	6 of 536 (1.1%)
			Sub: 38 of 536 (7.1%)
Switched interval			11 of 536 (2.1%)
Total	90 of 536 (16.8%)	67 of 536 (12.5%)	49 of 536 (9.1%)

Switched interval refers to neurons whose preferred interval changed from the sample presentation in one category scheme to the memory delay in the other category scheme or vice versa.

category scheme. By contrast, there was little difference in activity in response to images from the two different categories when the monkey grouped the same images under the other category scheme (Fig. 3*B*). We will refer to this as the neuron's "nonpreferred" category scheme. The insets in Figure 3, *A* and *B*, shows the neuron's average activity in response to each image (averaged across both the sample and delay intervals). Note that for the preferred category scheme, the neuron showed similar activity in response to the images from the same category and different activity in response to images from different categories (Fig. 3*A*). By contrast, there is little evidence for the neuron grouping the images by category under the nonpreferred scheme (Fig. 3*B*).

This category sensitivity under one but not the other category scheme was true for most PFC neurons (Table 1). Relatively few neurons (7.1% or 38 of 536) showed significant category sensitivity under both category schemes (13 neurons during the sample interval, 19 neurons during the memory delay, and 6 during both intervals; t test, $p < 0.05$, as above, Bonferroni corrected). This is about the level of overlap expected by chance and thus suggests independent representation of the categories in the PFC neuron population. Only 2.1% (11 of 536 neurons) showed significant category sensitivity for both category schemes, but for different schemes in the sample and delay intervals (t test, $p < 0.05$, as above, Bonferroni corrected). By contrast, many more neurons (29.3% or 157 of 536 neurons) showed significant category sensitivity for only one category scheme, but not both (53 neurons during the sample, 73 neurons during the memory delay, and 31 neurons during both intervals).

These neurons showed the same hallmarks of category representation that we found in previous studies (Freedman et al., 2001, 2002, 2003): a sharp transition in activity across the category boundary and greater selectivity across than within categories. Figure 4 plots the average activity of all PFC neurons with significant category sensitivity (t test, $p < 0.05$, as above) to images at different morph levels. On either side of the preferred category boundary, the average population activity is similar to images at each morph level. However, there is a sharp change in activity across the category boundary. For the nonpreferred category scheme, there is no sharp transition in activity across the category boundary; average activity is similar for all morph levels and across both categories. In other words, little category selectivity is apparent for the nonpreferred category scheme.

To quantify category selectivity, we used a category index that we have used in prior work (Freedman et al., 2001, 2002). It was calculated for each neuron with significant category sensitivity (t test, as above) using each neuron's difference in average activity in response to pairs of images from the same category (WCD) and to pairs of images from different categories (BCD), using images from the morph lines that crossed the category boundary. A standard index was computed for each neuron by dividing the difference between its BCD and WCD values by their sum. Posi-

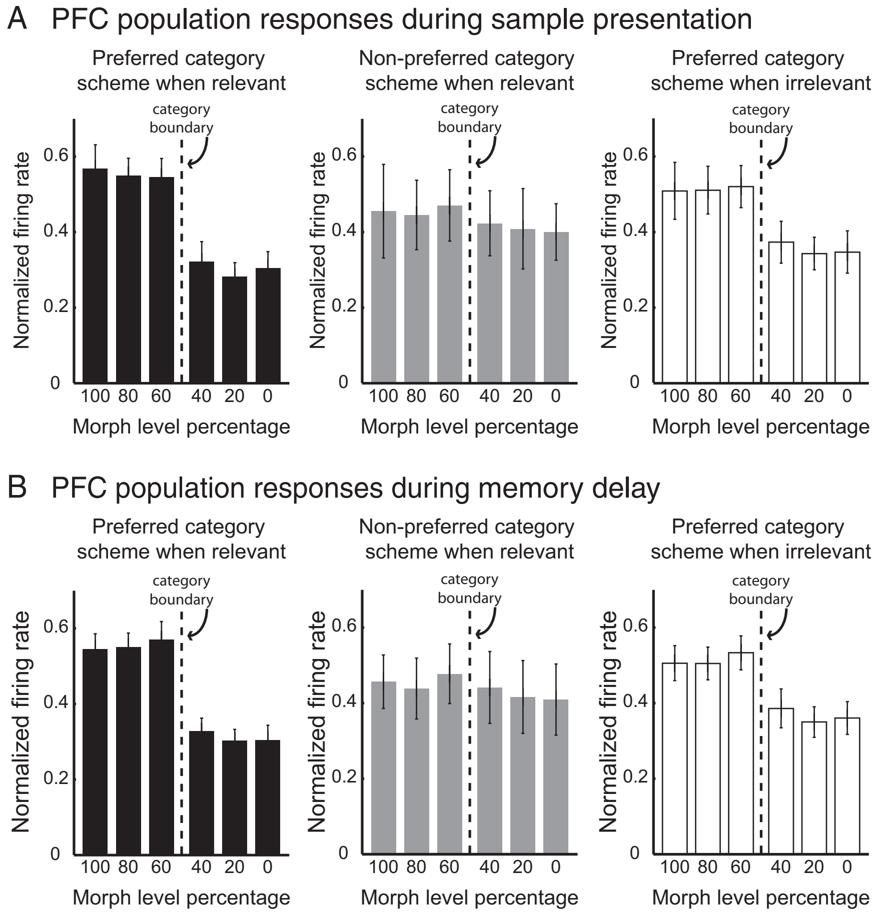

Figure 4. The category effect in average population activity. ***A***, ***B***, Normalized activity of PFC neurons to the images in each morph level collapsed across the six morph lines, sorted by preferred category scheme when relevant (left panel), nonpreferred scheme when relevant (center panel), and preferred scheme when irrelevant (right panel) during the sample presentation (***A***) and memory delay (***B***).

tive values indicate a larger difference between categories (i.e., a category effect) whereas negative values reflect larger differences within a category than between categories.

Figure 5*A* shows the distribution of category indices for each neuron's preferred category when it was relevant. The distribution was shifted significantly positive, i.e., there was a significant category effect (sample presentation mean index = 0.10, $p = 3 \times 10^{-5}$, *t* test vs mean of zero; memory delay = 0.16, $p = 1 \times 10^{-15}$, *t* test vs mean of zero). When the monkeys were performing each neuron's nonpreferred category scheme, category index values were significantly smaller (sample interval mean = 0.03, *t* test vs preferred scheme when relevant, $p = 0.002$; memory delay mean = 0.08, *t* test vs preferred scheme when relevant, $p = 0.002$) (Fig. 5*B*). Table 2 shows the population mean category index values across both monkeys (reported above) as well as the mean index values for each monkey individually. It shows that similar category effects were seen in each monkey.

Category effects were modified by category relevance

Because the same images were used in both category schemes, we were able to group the images according to a given category scheme both when that scheme was behaviorally relevant and when it was irrelevant (because the monkeys were performing the other scheme). The neuron in Figure 3 showed a smaller difference in activity between the categories of the preferred category scheme when it was irrelevant (Fig. 3*C*) compared with when it was relevant (Fig. 3*A*). Figure 4 shows the average activity across all the neurons that showed significant category sensitivity to their preferred category scheme (left panels) (*t* test, as above). The right panels show activity in response to the neuron's preferred category scheme when it was irrelevant. The sharp transition across the category boundary is still evident but the difference in activity in response to images from the two categories is smaller than when the (preferred) category scheme is relevant. In fact, the mean category index was significantly lower for the preferred category scheme when it was irrelevant (sample interval mean = 0.04, memory delay mean = 0.08) (Fig. 5*C*) than when it was relevant (sample interval mean = 0.10, memory delay mean = 0.16; *t* tests, preferred scheme relevant versus irrelevant, both $p < 0.01$) (Fig. 5*A*). Table 2 lists both the mean index values across both monkeys and for each monkey individually; similar effects were seen in each monkey.

This decrease in category selectivity when the preferred category scheme was irrelevant was due to a decrease in selectivity per se and not an overall decrease in neural activity. This is illustrated in Figure 5*D*, which plots the average activity of each neuron (across both categories) for both the sample presentation and memory delay intervals when the preferred category scheme was relevant against when it was irrelevant. There is a high degree of correlation between them ($R^2 = 0.81$) and the slope is nearly 1.0, indicating that overall neural activity is virtually the same when the monkeys were performing each neurons' preferred versus nonpreferred category scheme. Thus, as the single neuron example in Figure 3*C* implies, there was a loss of selectivity per se when each neuron's preferred category scheme was irrelevant, not a general decrease in neural activity.

Category correlation matrix

To further illustrate these effects, we computed pairwise correlations between each neuron's average activity in response to a given image and all other images (see Materials and Methods). We used activity from the sample presentation and/or memory delay intervals (considered independently) from every neuron that showed significant category sensitivity in that interval (*t* test as above, $p < 0.05$, Bonferroni corrected). To compute the correlation matrix, all the images were lined up along each axis in the same order for a given category scheme such that images 1–14 were always from one category and images 15–28 were always from the other category. We then computed the correlation between each neuron's average activity in response to a given image and the other images for both the sample and delay intervals, obtaining a correlation value across the neuron population for each pair of images. High correlation values meant that the activity was relatively similar for the two images; low correlation values meant that activity was relatively dissimilar. All three main

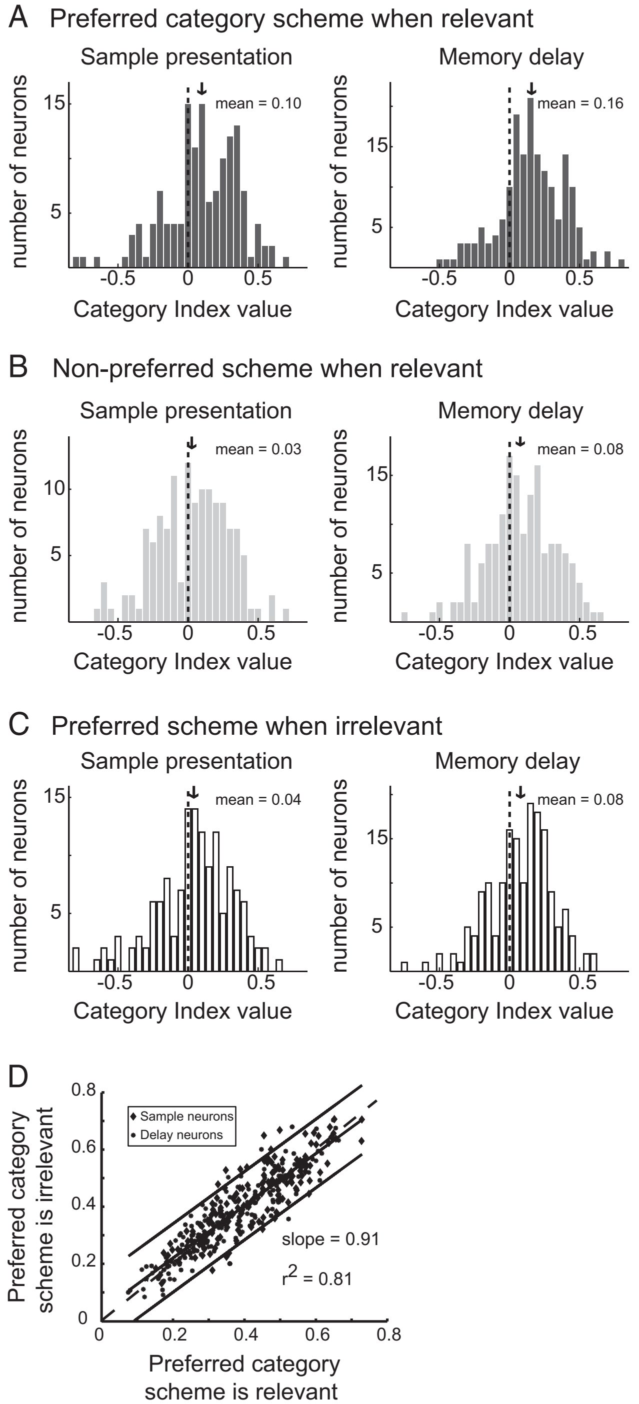

Figure 5. Distribution of category selectivity index values. ***A***, Index values of PFC neurons when the preferred category scheme was relevant during the sample presentation (left panel) and memory delay (right panel). ***B***, Index values for the same neurons when the nonpreferred scheme was relevant during the sample presentation (left panel) and memory delay (right panel). ***C***, Finally, index values for the preferred category scheme when irrelevant during sample (left panel) and memory delay (right panel). ***D***, Comparison of overall firing rate of each category-sensitive neuron for the preferred scheme and when the preferred scheme was irrelevant for sample presentation (◆) and memory delay-sensitive neurons (●).

results discussed above can be seen in these correlation matrices (Fig. 5).

Figure 6*A* shows the category effect. It plots the correlation coefficients for pairwise image comparisons for the preferred category scheme of each neuron when that scheme was relevant during the sample presentation (left panel) and memory delay (center panel). In the right panel is a map of the comparisons. Four large boxes corresponding to each quadrant are apparent in Figure 6*A*, left and right panels. They indicate higher correlation values (more similar activity) between images from the same category (upper right and lower left quadrants) than for comparisons between images from different categories (lower left and upper right quadrants). The zone marked "Same Category" indicates correlations between images that were members of the same category, but not physically similar because they were derived from different prototypes. Their correlation values (mean ± SEM: sample presentation = 0.21 ± 0.01, $p = 2 \times 10^{-4}$, 5000 permutation tests; memory delay = 0.26 ± 0.008, $p = 2 \times 10^{-4}$, 5000 permutation tests) were significantly higher than correlations between images belonging to different categories ("Different Categories," mean ± SEM: sample presentation = −0.20 ± 0.008, $p = 2 \times 10^{-4}$, 5000 permutation tests; *t* test, $p = 9 \times 10^{-95}$; memory delay = −0.17 ± 0.006, $p = 2 \times 10^{-4}$, 5000 permutation tests; *t* test, $p = 4 \times 10^{-130}$). The highest correlations (most similar activity) were between physically similar images from the same category and derived from the same prototype ("Same Prototype and Category," mean ± SEM: sample presentation = 0.39 ± 0.01, $p = 2 \times 10^{-4}$, 5000 permutation tests; memory delay = 0.38 ± 0.01, $p = 2 \times 10^{-4}$, 5000 permutation tests). Thus, in the PFC, there is a mixture of signals about physical appearance and category membership.

Figure 6*B* illustrates the category effects when each neuron's nonpreferred category scheme was relevant. During both the sample presentation (left panel) and memory delay (right panel), there were still relatively high correlation values for physically similar images (Same Prototype and Category, sample presentation = 0.31 ± 0.01, $p = 2 \times 10^{-4}$, 5000 permutation tests; memory delay = 0.32 ± 0.01, $p = 2 \times 10^{-4}$, 5000 permutation tests). But now there are boxes of relatively high correlations off the diagonal identity line in the Different Categories zones (sample presentation = 0.1 ± 0.01, $p = 6 \times 10^{-4}$, 5000 permutation tests; memory delay = 0.11 ± 0.007, $p = 2 \times 10^{-4}$, 5000 permutation tests). These are the effects of the other preferred category "leaking through," that is, still having an effect on activity when the nonpreferred category scheme is relevant.

Figure 6*C* illustrates the same comparisons for the same neurons as Figure 6*A*, but now it shows the effects of each neuron's preferred category when the nonpreferred category scheme was relevant. The effects were weaker in general, but more so for category per se than physical similarity (Same Category vs Same Prototype and Category). This is illustrated explicitly in Figure 6*D*, which plots the average correlation values for Same Category comparisons versus the Same Prototype and Category comparisons as a function of whether the preferred category scheme was relevant (Fig. 6*A*) or irrelevant (Fig. 6*C*). Average correlation values for each decreased when the category scheme was irrelevant, but there was a greater decrease for the Same Category correlations (sample presentation relevant = 0.21 ± 0.01, $p = 2 \times 10^{-4}$, 5000 permutation tests; irrelevant = 0.1 ± 0.01, $p = 0.003$, 5000 permutation tests; memory delay relevant = 0.26 ± 0.008, $p = 2 \times 10^{-4}$, 5000 permutation tests; irrelevant = 0.11 ± 0.007, $p = 0.009$, 5000 permutation tests) than for the Same Prototype and Category correlations (sample presentation relevant = 0.39 ± 0.01, $p = 2 \times 10^{-4}$, 5000 permutation tests; irrelevant = 0.31 ± 0.01, $p = 2 \times 10^{-4}$, 5000 permutation tests; memory delay relevant = 0.38 ± 0.01, $p = 2 \times 10^{-4}$, 5000 permutation tests; irrelevant = 0.32 ± 0.01, $p = 2 \times 10^{-4}$, 5000 permutation tests). This was confirmed by a two-way ANOVA that used category

Table 2. Category index values across both monkeys and for each monkey individually

	Both monkeys		Monkey O		Monkey L	
	Average index	*p* value against zero	Average index	*p* value against zero	Average index	*p* value against zero
Sample presentation						
Pref rel	0.10	3×10^{-5}	0.16	7×10^{-6}	0.07	0.01
Non-pref rel	0.03	0.24	−0.01	0.71	0.05	0.05
Pref irrel	0.04	0.08	0.05	0.10	0.04	0.11
Memory delay						
Pref rel	0.16	1×10^{-15}	0.16	3×10^{-9}	0.16	3×10^{-8}
Non-pref rel	0.08	1×10^{-4}	0.07	0.01	0.08	1×10^{-3}
Pref irrel	0.08	2×10^{-5}	0.08	1×10^{-3}	0.08	1×10^{-3}

p values show results of *t* test against a mean index value of zero (i.e., no category effect). Pref rel, Preferred category scheme when relevant; Non-pref rel, nonpreferred category when relevant; Pref irrel, preferred category scheme when irrelevant.

type (Same Category vs Same Prototype and Category) and category relevance as factors (Fig. 6*D*). It revealed a main effect of category type (average of sample presentation and memory delay, $p = 1 \times 10^{-4}$) and a main effect of relevance (average of sample presentation and memory delay, $p = 1 \times 10^{-4}$) as well as a significant interaction between the factors (average of sample presentation and memory delay, $p = 4 \times 10^{-5}$), indicating a greater effect of relevance on the Same Category comparisons than the Same Prototype and Category comparisons. In other words, when each PFC neuron's preferred category scheme is relevant (Fig. 6*A*), their activity reflects category membership and physical appearance. But when the other category scheme is relevant (Fig. 6*C*), information about the preferred category scheme weakens more than information about physical appearance. Thus, not only do PFC neurons tend to represent one, but not the other, category scheme, they still convey weak information about their preferred category scheme when it is not relevant.

To better visualize these effects, an MDS analysis was applied to pairwise correlation data presented above. The MDS reduces the dimensionality to determine whether some of the data clusters together in multidimensional space. Figure 7*A* illustrates the responses of PFC neurons when the preferred category scheme was relevant. It shows that the images generally clustered into two groups, each corresponding to the two categories of each neuron's preferred scheme. The Same Prototype and Category images are denoted by the dashed lines. Note that numbers in Figure 7 correspond to the ordering of images within a category scheme, not the exact image. There is still some separation based on physical similarity; images from the same prototype tend to cluster together (images 1–7, 8–14, 15–21, and 22–28). The prototype clusters from the same category are near each other and are distinct from that of the other category. In the memory delay (Fig. 7*A*, middle panel), the prototype clusters overlap, suggesting that the category membership is more important that the physical similarity. In fact, the mean Euclidean distance (mean ± SD) between images within the same category (sample presentation images 1–14: 0.32 ± 0.29; images 15–28: 0.31 ± 0.27; average within category: 0.32 ± 0.28 and memory delay images 1–14: 0.27 ± 0.23; images 15–28: 0.28 ± 0.25; average within category: 0.28 ± 0.24) were significantly smaller than the mean distances between images from different categories (sample presentation: 0.88 ± 0.52; $p = 1 \times 10^{-4}$; memory delay: 0.86 ± 0.52; $p = 1 \times 10^{-4}$). The first two dimensions were used in the calculation of the Euclidean distances because they accounted for 60.9% of the variance of the sample presentation and 56.4% of the variance of the memory delay (Fig. 7*A*, right panel). The addition of a third dimension only explained 7–8% more of the variance.

Figure 7*B* shows the data when the images were sorted by each neuron's nonpreferred category scheme when it was relevant. Now, the spatial arrangement of the images derived from the different prototypes changes. Images from the same (nonpreferred) category no longer localize overlapping or near one another (average WCD sample presentation: 0.5 ± 0.44 vs 0.32 ± 0.28 of the preferred scheme when relevant, *t* test $p = 2.5 \times 10^{-11}$; memory delay: 0.46 ± 0.46 vs 0.28 ± 0.24 of the preferred scheme when relevant, *t* test $p = 1.6 \times 10^{-11}$). Instead, the clusters that tend to be nearer one another are derived from prototypes from the same category under the preferred category scheme. The first two dimensions were used in the calculation of the Euclidean distances because they accounted for 52.9% of the variance of the sample presentation and 57.7% of the variance of the memory delay (Fig. 7*B*, right panel). Once again, it illustrates the effects of each neuron's preferred category scheme affecting neural activity even when the monkeys are performing the other scheme.

As predicted from Figure 6*C*, when the preferred category scheme was irrelevant (Fig. 7*C*), the clustering by category was diminished. The mean Euclidean distance for images from the same category under the preferred scheme are significantly larger when the scheme was irrelevant (sample presentation: 0.37 ± 0.31; memory delay: 0.34 ± 0.30) (Fig. 7*C*, left panel) compared with when that preferred scheme was relevant (sample presentation: 0.32 ± 0.28, *t* test, $p = 1 \times 10^{-2}$; memory delay: 0.28 ± 0.24, *t* test, $p = 1.1 \times 10^{-3}$) (Fig. 7*C*, middle panel). Further, the distance between images from the same category but different prototypes (images 1–7 vs 8–14, images 15–21 vs 22–28) also increases when the preferred category scheme is irrelevant (sample presentation images 1–7 vs images 8–14: irrelevant, 0.57 ± 0.43; relevant, 0.46 ± 0.33; *t* test $p = 0.04$; images 15–21 vs images 22–28: irrelevant, 0.39 ± 0.31; relevant, 0.29 ± 0.29; $p = 0.01$; memory delay images 1–7 vs images 8–14: irrelevant, 0.45 ± 0.34; relevant, 0.24 ± 0.24; *t* test $p = 9 \times 10^{-7}$; images 15–21 vs images 22–28: irrelevant, 0.47 ± 0.34; relevant, 0.34 ± 0.30, $p = 0.03$). This greater difference between images derived from the same prototype further supports the conclusion that there is increased encoding of physical similarity by neurons when their preferred category scheme is irrelevant.

Category and match/nonmatch effects

At the end of the trial, a test image was presented and monkeys had to judge whether it matched the category of the sample. As in our prior studies (Freedman et al., 2001, 2002, 2003), we found that many PFC neurons modified their activity depending on the test image category match/nonmatch status. Of the entire population of recorded neurons, 72 of 536 (13.4%) were significantly sensitive to whether the test stimulus was a category match or a nonmatch (*t* test of all match trials compared with all nonmatch trials, $p < 0.01$). Thirteen neurons showed greater activity for matches and 59 neurons showed greater activity for nonmatches.

To test the effects of the current category scheme on the match/nonmatch effects, a two-way ANOVA (evaluated at $p < 0.01$) was calculated on the test stimulus activity for all category-sensitive neurons. One factor was whether the monkey was performing the neuron's preferred or nonpreferred category scheme and the other factor was whether the test stimulus was a category match or a nonmatch to the sample. The results are summarized in Table 3. Approximately 32% (66 of 206) of category-sensitive neurons showed a significant main effect and/or an interaction between the factors. Only 4.4% (9 of 206) showed a main effect of the preferred/nonpreferred category scheme, which is consistent with our other observation that overall activity of neurons is similar whether the preferred or nonpreferred category scheme is relevant. A larger proportion of neurons, 19.9% (41 of 206), showed a significant main effect of match/nonmatch. Of these 41 neurons, 29.3% (12 of 41) also showed a significant interaction with the preferred/nonpreferred category status (i.e., they showed stronger match/nonmatch effects under one of the category schemes). Most of them (10 of 12 or 83.3%) showed stronger effects of category matching when the monkey was performing the nonpreferred category scheme of the neuron. Approximately 7.8% (16 of 206) of neurons showed only an interaction effect between the preferred/nonpreferred category status and match/nonmatch. The majority (10 of 16 or 62.5%) showed stronger category match effects for the neuron's nonpreferred category. Together, these results demonstrate that some PFC neurons reflected the match/nonmatch status of the test stimuli and possibly the corresponding behavioral response (i.e., letting go of the lever on a match).

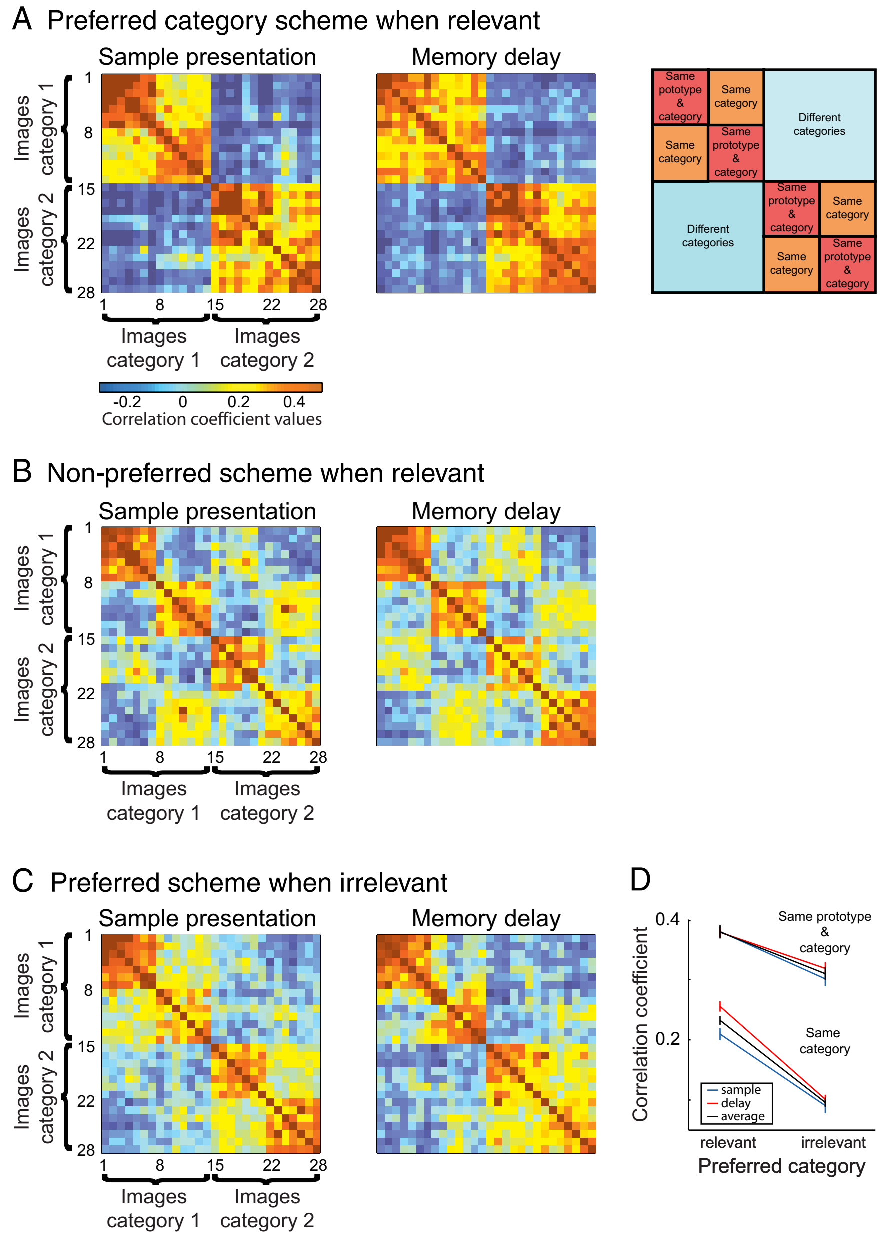

Figure 6. Category selectivity reflected in correlations between pairs of images during the sample presentation and memory delay. ***A***, Activity in response to the preferred category scheme when it was relevant. Neurons respond more similarly (higher correlations) to images from the same category. To the right is a map of the image comparisons. ***B***, The same neurons carried virtually no information about the nonpreferred category scheme. ***C***, Correlations were weaker to the preferred category scheme when it was irrelevant. ***D***, Comparison of average correlation values showed there was a greater effect of category relevance on category information per se.

Discussion

We report two main results. First, when monkeys switched between two orthogonal category schemes for the same image set, independent neuron populations in the PFC reflected the category distinctions under the different category schemes. Second, the effects of category membership were weakened, but not absent, when monkeys were performing an alternative, competing category distinction.

We could have found that neurons were multitaskers; that is, many neurons could have shown sensitivity for all four of the categories (2 categories × 2 schemes) in a graded fashion (each category eliciting a unique level of neural activity). Or neurons could have multitasked by modifying their sensitivity (i.e., changing which images elicit a given level of activity for each category scheme). Instead, we found that most PFC neurons only showed selectivity or much more selectivity for categories under one category scheme. These data seem to suggest a sparse encoding scheme with different neurons representing each category (Vinje and Gallant, 2000; Olshausen and Field, 2004). Together, these results support a model of human object recognition (Riesenhuber and Poggio, 2000, 2002) that predicts that different categorization tasks on the same stimuli are mediated by different neuron populations or circuits in PFC. Humans could learn our task much more quickly than monkeys, raising the possibility of dif-

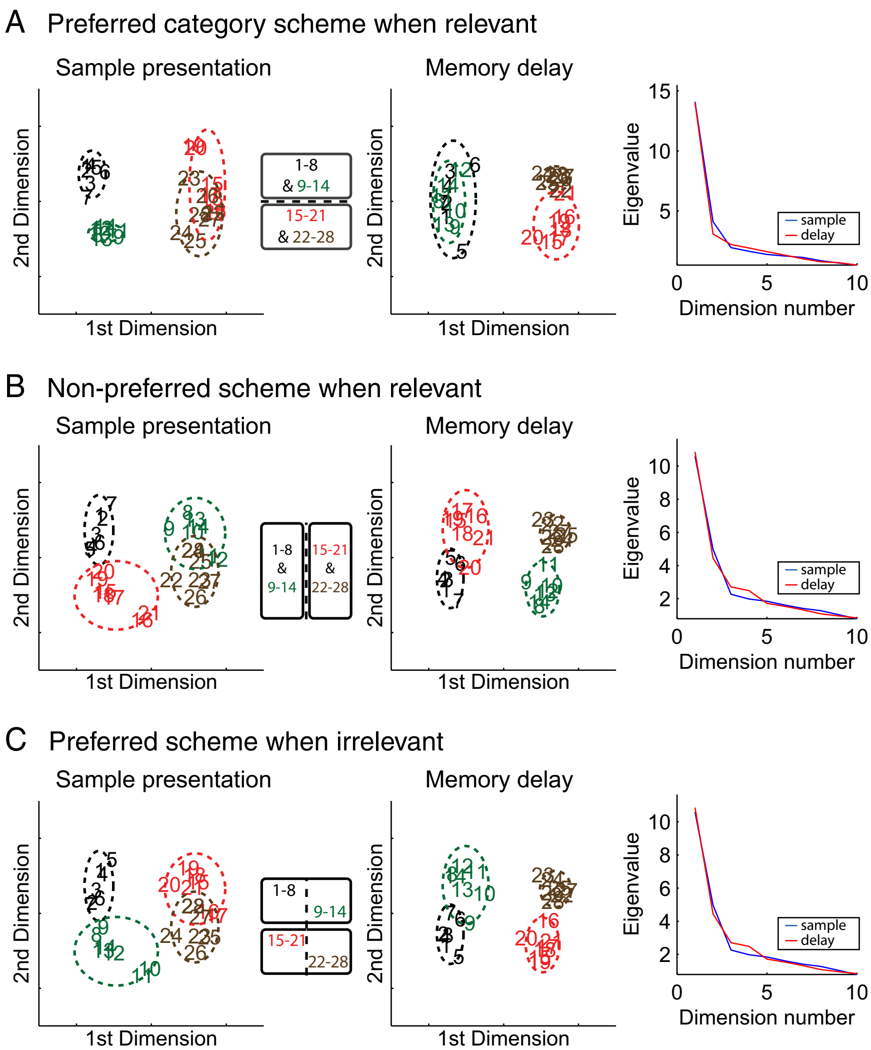

Figure 7. Multidimensional scaling of correlation coefficients during the sample presentation and memory delay. ***A***, The images tended to cluster into two groups corresponding to the two categories (images 1–14 and images 15–28) of the preferred category scheme. The representation of the constituent images (dashed lines) from the two prototypes of each category overlapped. ***B***, The clustering conveyed little information about the nonpreferred category scheme when it was relevant. ***C***, When the preferred category scheme was irrelevant, the clustering by category diminished and clustering of same prototype and category images increased. The scree plots (right panels) indicate that the first two dimensions were sufficient to describe the data. The legends show where the images were in stimulus space for each category scheme.

Table 3. Category-sensitive PFC neuronal responses to test image presentation

Test image presentation	No. of category-sensitive neurons
Match/nonmatch effect	41 of 206 (19.9%)
With an interaction effect	12 of 41 (29.3%)
M/NM effect greater for preferred	2 of 12 (16.7%)
Nonpreferred	10 of 12 (83.3%)
Preferred/nonpreferred effect	9 of 206 (4.4%)
Preferred	5 of 9 (55.6%)
Nonpreferred	4 of 9 (44.4%)
Interaction-only effect	16 of 206 (7.8%)
M/NM effect greater for preferred	6 of 16 (37.5%)
Nonpreferred	10 of 16 (62.5%)

M/NM, Match/nonmatch.

ferent neural substrates between them. However, our results also agree with recent human neuroimaging findings that when humans were trained on a similar categorization task, circuits coding for the trained categorization scheme were suppressed when subjects were executing a different task on the same stimuli (Vogels et al., 2002; Jiang et al., 2007). In any case, although our monkeys did require a great deal of training, humans often have years of experience with familiar categories. Thus, our task likely tapped into the neural substrates that allow humans to flexibly categorize familiar categories. Finally, as in previous studies, many neurons showed category match/nonmatch effects, i.e., different levels of activity in response to a test stimulus depending on whether it matches the category of the sample. This suggests that this population of neurons may also contribute to the category match judgments in addition to representing the categories per se. A significant proportion of them showed stronger match/nonmatch effects for the neuron's nonpreferred category scheme, suggesting some independence (or at least a nonstraightforward relationship) between the neuron ensembles that represent the category and those that determine category matching.

Our results seems to contrast with studies and theories suggesting that many PFC neurons multitask (Duncan and Miller, 2002). However, the independence of the neural representation of the two category schemes may have been due to the demands of our task: the two different category schemes were in direct competition. Because images had to be categorized differently under the two different category schemes, there was a high probability of miscategorization by the wrong scheme. Such interference has been observed when humans switch between different category tasks; images from one task are often miscategorized in the other (VanRullen and Thorpe, 2001). Thus, the brain may have reduced the chances of this error by representing the category distinctions in two independent neuron populations, rather than multiplex the representations onto overlapping populations of neurons. The lingering effects of the neurons' preferred category scheme when the nonpreferred scheme was being performed underscored the utility of their independent representation for reducing errors, although it is possible that this effect could have been due, in part, to the monkeys thinking they were performing the currently irrelevant category scheme and accidentally responding correctly under the relevant scheme. Or alternatively, the monkeys could be trying to solve the task with the schemes weighted probabilistically on each trial (e.g., they could have given the relevant scheme 80% weight and the irrelevant scheme 20%) instead of in a binary fashion. The monkeys' high level of performance suggests that the contribution of these alternatives would be minimal. It is also possible that the categories were represented in independent neurons because the monkeys learned the categories separately before learning to flexibly switch between them. In any case, the weakening of selectivity for the preferred category scheme when it was irrelevant seems to suggest a mechanism for suppressing irrelevant competing information. One could test these possibilities by training monkeys on independent, noncompeting category distinctions. If, by contrast, neurons multitask these categories, it would suggest the distribution of information across PFC neuron ensembles may be arbitrary. In other words, it may be the that the nature and organization of PFC ensembles are far more dependent on top–down behavioral demands than bottom–up sensory inputs, a sharp contrast to the organization seen in sensory cortex.

References

DeGutis J, D'Esposito M (2007) Distinct mechanisms in visual category learning. Cogn Affect Behav Neurosci 7:251–259.

Diester I, Nieder A (2007) Semantic associations between signs and numerical categories in the prefrontal cortex. PLoS Biol 5:e294.

Duncan J, Miller EK (2002) Cognitive focusing through adaptive neural coding in the primate prefrontal cortex. In: Principles of frontal lobe function (Stuss D, Knight RT, eds), pp 278–291. New York: Oxford UP.

Freedman DJ, Assad JA (2006) Experience-dependent representation of visual categories in parietal cortex. Nature 443:85–88.

Freedman DJ, Miller EK (2008) Neural mechanisms of visual categorization: insights from neurophysiology. Neurosci Biobehav Rev 32:311–329.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291:312–316.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2002) Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. J Neurophysiol 88:929–941.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J Neurosci 23:5235–5246.

Fuster JM (2000) Prefrontal neurons in networks of executive memory. Brain Res Bull 52:331–336.

Goldman-Rakic P (1987) Circuitry of prefrontal cortex and regulation of behavior by representational memory. In: Handbook of physiology: the nervous system: higher functions of the brain, pp 373–417. Bethesda, MD: American Physiological Society.

Hegdé J, Van Essen DC (2006) Temporal dynamics of 2D and 3D shape representation in macaque visual area V4. Vis Neurosci 23:749–763.

Hill EL (2004) Executive dysfunction in autism. Trends Cogn Sci 8:26–32.

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

Kéri S, Szekeres G, Szendi I, Antal A, Kovács Z, Janka Z, Benedek G (1999) Category learning and perceptual categorization in schizophrenia. Schizophr Bull 25:593–600.

Manly BFJ (1997) Randomization, bootstrap and Monte Carlo methods in biology. London: Chapman and Hall.

Miller EK (2000) The prefrontal cortex and cognitive control. Nat Rev Neurosci 1:59–65.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.

Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. Curr Opin Neurobiol 14:481–487.

Paxinos G, Huang XF, Toga AW (2000) The rhesus monkey brain in stereotaxic coordinates. San Diego: Academic.

Poldrack RA, Rodriguez P (2004) How do memory systems interact?: Evidence from human classification learning. Neurobiol Learn Mem 82:324–332.

Riesenhuber M, Poggio T (2000) Models of object recognition. Nat Neurosci 3 [Suppl]:S1199–S1204.

Riesenhuber M, Poggio T (2002) Neural mechanisms of object recognition. Curr Opin Neurobiol 12:162–168.

Shelton C (2000) Morphable surface models. Int J Comp Vis 38:75–91.

Tan HY, Sust S, Buckholtz JW, Mattay VS, Meyer-Lindenberg A, Egan MF, Weinberger DR, Callicott JH (2006) Dysfunctional prefrontal regional specialization and compensation in schizophrenia. Am J Psychiatry 163:1969–1977.

Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323.

VanRullen R, Thorpe SJ (2001) Is it a bird? Is it a plane?: Ultra-rapid visual categorisation of natural and artifactual objects. Perception 30:655–668.

Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273–1276.

Vogels R (1999) Categorization of complex visual images by rhesus monkeys. Part 2. Single-cell study. Eur J Neurosci 11:1239–1255.

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

Weickert TW, Goldberg TE, Callicott JH, Chen Q, Apud JA, Das S, Zoltick BJ, Egan MF, Meeter M, Myers C, Gluck MA, Weinberger DR, Mattay VS (2009) Neural correlates of probabilistic category learning in patients with schizophrenia. J Neurosci 29:1244–1254.