SUPPLEMENTARY INFORMATION

This file contains **Supplementary Methods**, Supplementary Figures 1-6 accompanied by Legends, and Supplementary Data. **Figure 1** is for a schematic of the main finding, and **Figure 2** is for illustrating the behavioral task. The supplementary **Figures 3 and 4** collectively show that the variations of preparatory activity of each PF cell for sequences belonging to one category were small, whereas the variations for sequences belonging to different categories were large. The supplementary **Figure 5** shows how the category selective activity developed during performance of successive trials. The supplementary **Figure 6** presents the category selectivity during the time epoch preceding the preparatory period. Finally, **the supplementary data** includes information about (1) the behavioral data for error trials, (2) results of analysis on 21 cells during performance of error trials, (3) the comparison of data obtained in dorsal and ventral PF, and (4) the category selectivity during the inter-trial interval.

SUPPLEMENTARY METHODS

Behavioral task and recording methods. We trained two monkeys (Macaca fuscata) to perform a series of four movements with intervening intervals of 1.0-1.4 s in eleven different temporal orders (see Supplementary Figure 2). Each movement was either a push, a pull, or a turn of a handle, which was grasped with the right hand. After an inter-trial interval of 1.5 s, the task began when the monkey held the handle in a neutral position and waited for a period of 2.5 to 4 s, whereupon a tone signal was presented to trigger the first of the four movements. After completion of an individual movement, a mechanical device returned the handle to a neutral position; the animal had to hold the handle in this position and wait for the next movement-trigger signal. Initially, during the learning phase, green, yellow, and red LEDs indicated that the monkey was required to push, pull, and turn the handle, respectively. The animals underwent five learning trials and subsequently performed the

sequential motor task in the absence of the visual signals. After five trials based on memory, the sequence of the movements was changed for the next set of trials. For the temporal structures of the 11 motor sequences, the four **paired** sequences consisted of two movement pairs, four **alternate** sequences consisted of alternation of two movements, and the three remaining **four-repeat** sequences required the one movement to be repeated four times. Thus, after the learning trials, monkeys had to prepare to perform a sequence of four movements based on memorized information about the temporal patterns that fell into one of three categories. A reward of fruit juice was delivered when the monkey accomplished the four movements in a correct sequence (0.5 s after the execution of the fourth movement). When 5 trials were accomplished under memory, a new sequence, selected unpredictably from the 11 sequences, commenced for the next visually-guided trials. We used conventional electrophysiological techniques to obtain in vivo single-cell recordings[10] from the lateral prefrontal cortex above and below the principal sulcus in the left hemisphere. Cortical sulci and recording locations were identified using a magnetic resonance imaging scanner before recording, and were verified by histological examination of Nissl-stained brain sections. We sampled cellular activity from the dorsolateral prefrontal cortex rostral to the frontal eye field, including the banks of the principal sulcus. We also monitored eye positions and velocity with an infrared corneal reflection monitoring system, and activity in the limb and trunk muscles with electromyography. Recordings were made from the following muscles: extensor digitorum communis, flexor digitorum profundus, exensor carpi ulnaris and radialis, flexor carpi radialis, biceps and triceps brachii, brachioradialis, deltoideus, sternomastoideus, trapezius, supraspinatus, pectoralis major, thoracic and lumbar paravertebral, iliopsoas, and quadriceps.

Training procedure. Both monkeys were first trained to perform one of three movements in response to one of three visual signals. A tone signal was given simultaneously with the visual signal. They learned to associate each color of the visual signal to each movement. Thereafter, the monkeys

were trained to perform one of the three movements four times under visual guidance, and then repeat the same movement four times without visual guidance. In that case, only the tone signal was given as a movement trigger. The next stage of training was the introduction of movement pairs (e.g., push and then pull) with an interval of 1 - 1.4 s. Initially, a pair of two movements were performed under visual guidance. After performing the pair of movements 5 times, the same pair had to be performed based on memory, in response to a tone signal. Subsequently, the training included a series of three movements in different sequences. When this was learned, the training reached the final stage of a sequence of four movements, initially under visual guidance and then based on memory. It took 12 months for both monkeys to learn the behavioral task including 11 sequences, and we kept training another 6 months until the task performance became stable with small error rates (1.8 – 2.3 % for both monkeys). Single-cell recordings commenced thereafter. During the entire period of training, no external signals were used to inform the monkeys a particular category of sequences. However, in the course of training, to facilitate monkey's learning, one particular category was, at times, taught repeatedly for tens of trials before another category was taught. This might have encouraged the monkeys to learn sequences as belonging to categories.

Error rates. We calculated the rate of occurrences of errors for each category of sequences during performance of the behavioral task when cellular activity was recorded. For the first monkey, the error rates were 1.8, 1.9, and 0.2 % for the categories of alternate, paired, and 4-repeat sequences, respectively. Error rates for the second monkey were 2.2, 2.3, and 0.5 % for the alternate, paired, and 4-repeat categories.

Data analysis. Our database included cells from which activity was recorded during at least two blocks of trials for each of eleven different sequences belonging to three categories with the monkeys performing based on visual cues or memory. We defined three periods for data analysis: the control period (500 ms during the initial hold period), the preparatory period (2 s before the first GO

signal), and the motor response period (from the start of the GO signal until completion of the fourth movement). Cellular activity was defined to be task-related if the discharge rate during either the preparatory or motor response periods differed significantly from that recorded during the control period (Wilcoxon signed-rank test; $\alpha = 0.05$). We focused on studying cells that had their activity modulated during the preparatory period. To statistically assess how the three categories of motor sequences were related to the activity of the cells, we performed linear regression analysis using the following regression model: firing rate $= \beta_0 + \beta_1 \times$ (categories). In this equation, β_0 is the intercept and β_1 is the regression coefficient. The categorical factor was the categories of sequences (paired, alternate, or four-repeat). The regressors were entered into the analysis as dummy variables. We calculated the probability (p value) that the coefficient equaled zero. If the $p<0.01$, we judged that the activity reflected the sequences of a category. Subsequently, for cells satisfying this criterion, we examined whether the cellular activity was related to an individual sequence of movements. For this purpose, we performed a second round of linear regression analysis using the following equation: firing rate $= \beta_0 + \beta_2 \times$ (sequences). In this equation, β_2 is the regression coefficient and the categorical factors for sequences were the 11 different movement sequences (see Fig. 1). The regressors were entered into the analysis as dummy variables. In this study, we defined the activity of a cell to be selectively reflecting a category if it was found to be related to the category but not to an individual sequence within this category. We also similarly analyzed eye positions and electromyography data with regression analysis. A/D converted data were used to perform the regression analysis. We found no statistically significant effects of the categories of behavioral sequences on these variables.

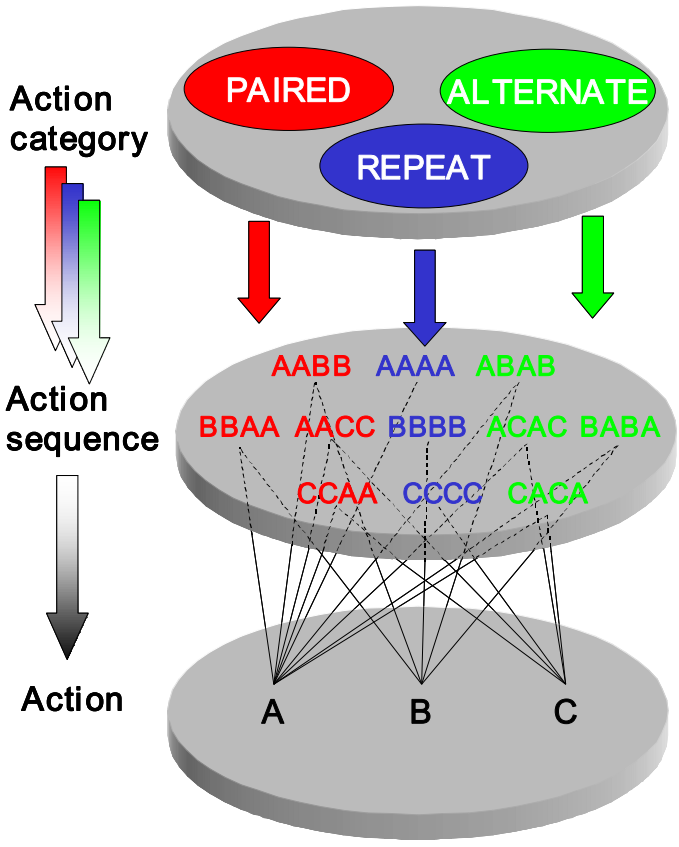
Furthermore, we determined an index that described the association between cellular activity during the preparatory period and the monkey's subsequent selection of a sequence from one of the three sequence categories. This index approximated the ability of an observer to predict the monkey's

behavior from the cellular activity. We performed receiver operating characteristic (ROC) analysis[12,13] to obtain this predictive index. For each cell, the two spike rate distributions for the preferred and non-preferred category were compared. To obtain the ROC curve, we first calculated the probability of true-positives based on the spike rate distribution for the preferred category and the probability of false-positives for the non-preferred category. Then, the probability of true-positives was plotted as a function of the probability of false-positives for a number of criteria. The area under the ROC curve was taken as a quantitative measure (ROC index) of how well the two distributions were separated, i.e., how well the activity of a cell discriminated between the preferred and non-preferred category. An index of 0.5 represents identical distributions (no discrimination for categories), whereas an index of 1.0 indicates completely separated distributions (perfect discrimination). This analysis was repeated for all three categories. The sliding ROC analysis (kernel width 500 ms, slid in 50-ms increments) was performed to derive a cell's index at each time point during the preparatory period.

For the purpose of displaying cellular activity during performance of the behavioral task, we used the following three display formats. (1) In the raster display (Figures 1 & 2), each dot shows when the cell discharged and each row of dots represents a trial with a particular sequence of movements, aligned on the onset of the GO signal for the first of memorized movements (i.e., at the end of the preparatory period). (2) Peri-event histograms show the sum of activity in 40-ms time bins. (3) To display the time courses of the activity of cell populations (Figure 2a) that exhibited selectivity for each category of sequences, spike density function was plotted for PF cells that were selectively active as the monkey prepared to perform behavioral sequences from one of the three categories. The spike density of each category-selective cell was calculated with a time bin of 150 ms and then averaged for each of motor sequences, after individually normalizing the spike densities to a maximal value during the preparatory period.

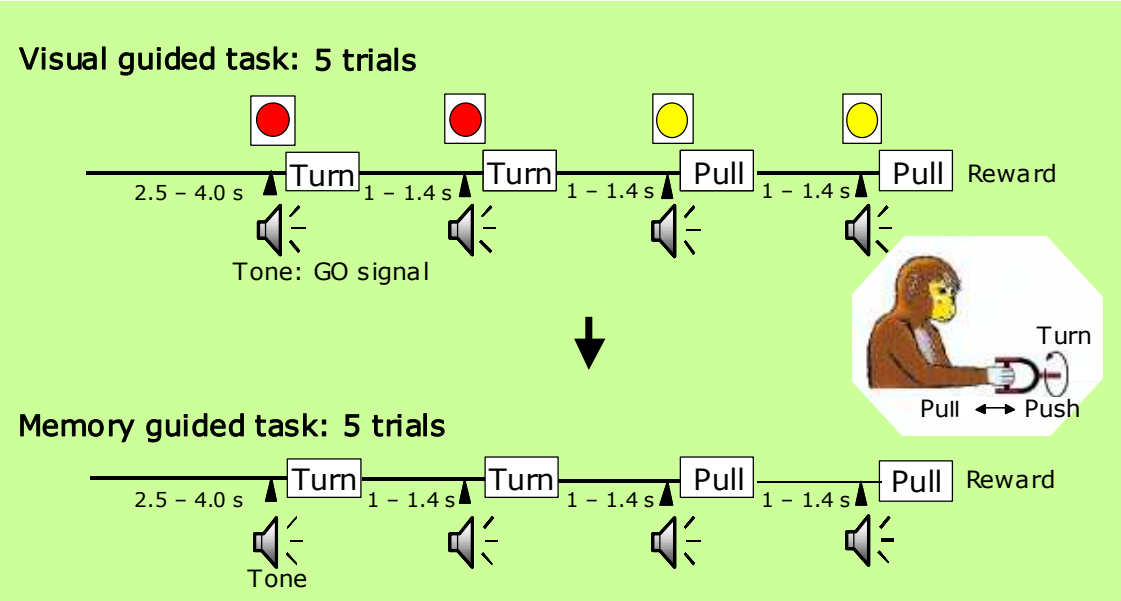
SUPPLEMENTARY FIGURES

Supplementary Figure 1



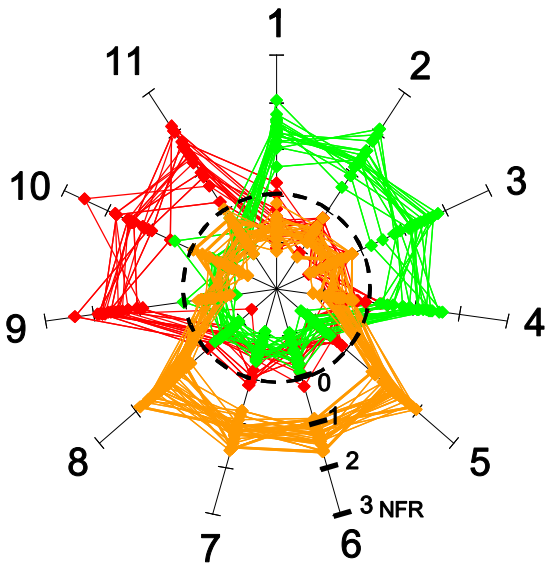
Supplementary Figure 1. A schematic drawing of the main finding in this study. Here we report that the category of actions is represented in the prefrontal cortex during behavioral planning. In daily life, actions are often performed with a particular temporal sequence that should be memorized and planned before execution. It has been reported previously that cortical cells, primarily in secondary motor areas, take part in planning individual action sequences (e.g., AABB, ABAB, etc.). Now we deal with the case when a subject has a large number of action sequences to be memorized and planned. In such a case, categorization of the sequences according to the specific temporal structure serves to facilitate memory-based planning. We found that the prefrontal cells represent such category of sequences as paired (e.g., AABB, CCAA), alternate (e.g., ABAB, CACA), or repeat (e.g., AAAA) types of sequences. Such categorization is a model for the conceptualization of macrostructure of behavioral plans.

Supplementary Figure 2



Supplementary Figure 2. A diagram illustrating the temporal sequence of events constituting the behavioral task used in this study. Initially, a correct sequence of 4 movements was guided with visual signals. During the performance of 5 trials under visual guidance, monkeys were required to remember that particular sequence, and then had to perform the same sequence based on memory. After the completion of 5 trials under memory guidance, the sequence of 4 movements was renewed in the subsequent visually guided trials. On performing both the visually guided and memory guided trials, monkeys had to wait 2.5-4 s while placing the handle in the neutral position. Each of the 4 movements, triggered with a tone signal, was temporally spaced with a variable interval of 1-1.4 s.

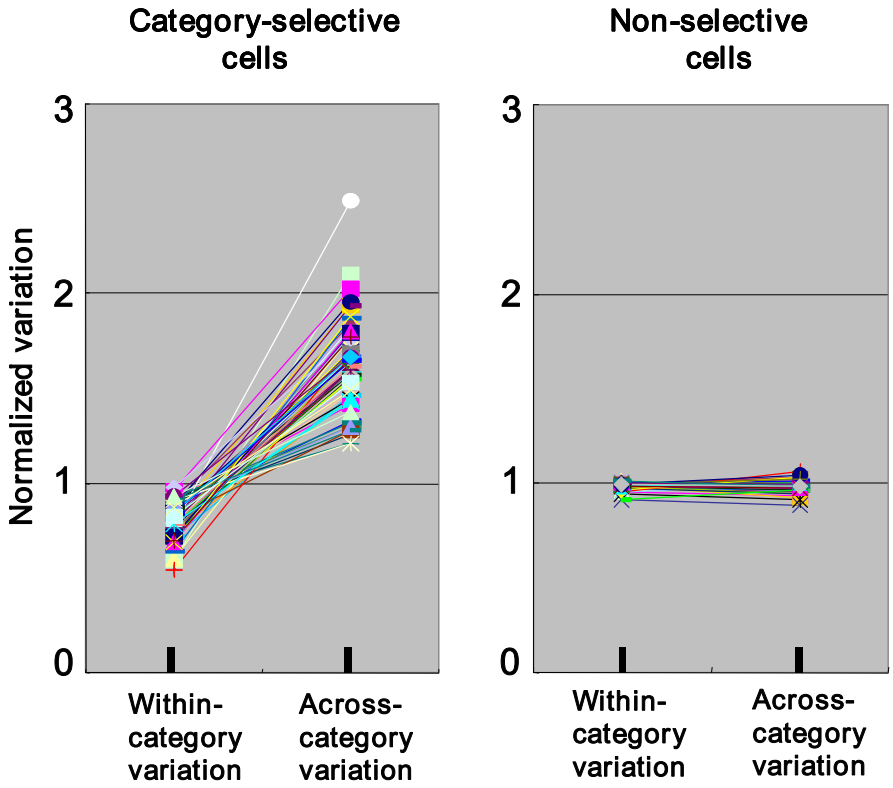
Supplementary Figure 3



paired sequences	alternate sequences	4-repeat sequences
1:Turn-Turn-Push-Push	5:Turn-Push-Turn-Push	9:Turn-Turn-Turn-Turn
2:Turn-Turn-Pull-Pull	6:Turn-Pull-Turn-Pull	10:Push-Push-Push-Push
3:Push-Push-Turn-Turn	7:Push-Turn-Push-Turn	11:Pull-Pull-Pull-Pull
4:Pull-Pull-Turn-Turn	8:Pull-Turn-Pull-Turn	

Supplementary Figure 3. A polar plot of the activity of 85 category-selective PF cells. Here we plot the activity of each cell during the preparatory period before the execution of each of the 11 motor sequences (last 500 ms in the preparatory period). The activity was normalized using the following formula: the normalized firing rate (NFR) = (the firing rate prior to the execution of each sequence – the mean firing rate before the execution of all the sequences)/the standard deviation of the firing rate. The NFRs calculated in all the trials were then averaged for each of the 11 sequences. The averaged NFR was plotted in a polar diagram in which the 11 axes correspond to the individual sequences and the radius of each axis denotes the mean frequency. Data for each cell selectively active before alternate, paired, and four-repeat sequences were performed are indicated in orange, green, and red, respectively. The dotted circle marks the zero level of the normalized firing rate.

Supplementary Figure 4



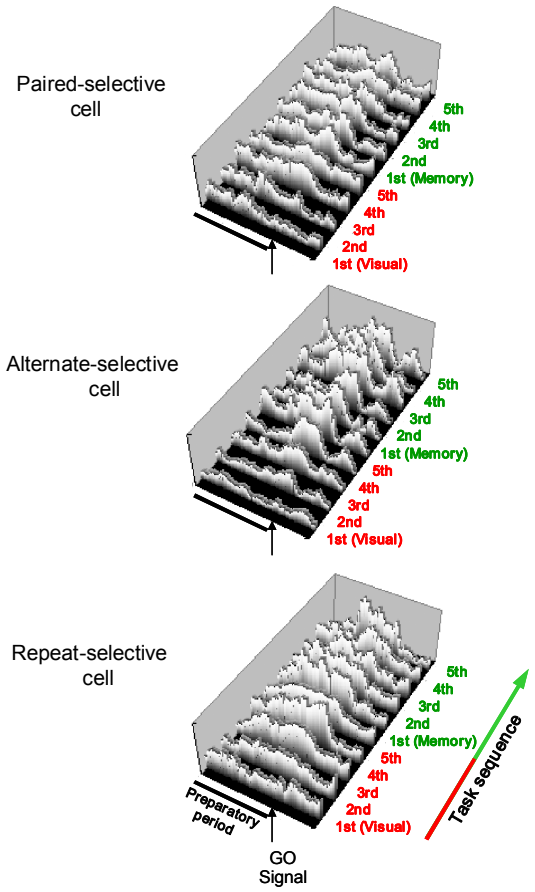
Supplementary Figure 4. The distribution of within-category and across-category variations. The variations in the preparatory activity of PF cells that were category selective were compared with the variations in the activity of nonselective cells. For each PF cell, we calculated the within-category variation (WCV) and the across-category variation (ACV) using the following equations:

$$WCV = SQRT(\sum_{i=1}^{N}(x_i - mean_a)^2 / N),$$

$$ACV = SQRT((\sum_{i=1}^{N}(x_i - mean_b)^2 + \sum_{i=1}^{N}(x_i - mean_c)^2)/2/N).$$

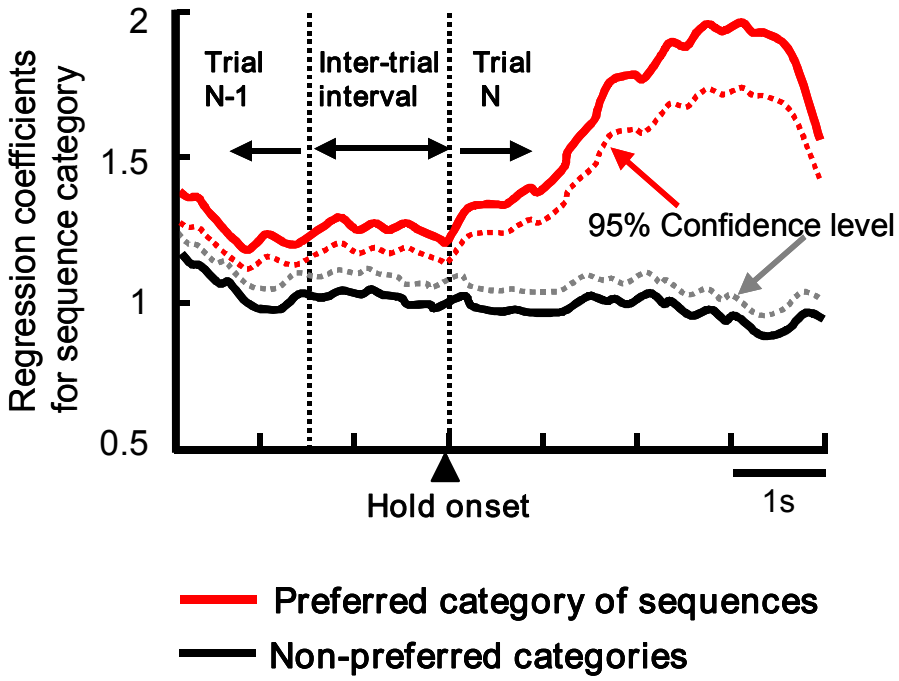
In these equations, x_i is the firing rate in the preparatory period in the ith trial, and $mean_a$ is the mean firing rate in the preparatory period before the monkey performs a sequence from category a to which the ith trial belongs. N is the number of trials. $Mean_b$ and $mean_c$ are the mean firing rates in the preparatory period before the monkey performs a sequence from categories b and c, respectively. For the category-selective PF cells, the ACVs were significantly greater than the WCVs ($p < 0.001$ by t-test). In contrast, for nonselective cells, the ACVs and WCVs were not different.

Supplementary Figure 5



Supplementary Figure 5. Time-varying plots of selectivity for the category of sequences in three PF cells developing during 5 visually-guided and 5 memory-guided trials. As the measure of the category selectivity, regression coefficients calculated by performing the regression analysis (looking for the relation of cellular activity to the category, as explained under methods), are taken on the ordinate. The regression coefficients are plotted for the time period starting from 4,000 ms before the GO signal and ending at 2,000 ms after the GO signal, with a time bin of 100ms. The analysis was performed for every trial from the first through the fifth trial under visual guidance, and under memory guidance. Three typical examples of category selective cells are illustrated from the top to the bottom. Note that, for each cell, the selectivity for the preferred category started to grow during early visually-guided trials, and grew successively to reach to a level exhibited steadily during memorized trials.

Supplementary Figure 6



Supplementary Figure 6. Category selectivity during the time epoch preceding the preparatory period. For 33 PF cells that exhibited category selectivity during the inter-trial interval, regression coefficients for the sequence category were calculated with the regression analysis explained in Supplementary Methods. To compare the regression coefficients calculated for preferred category of sequences vs. non-preferred categories at a population level, mean values of coefficients are plotted with the time bin of 100 ms, along with values indicating 95 % confidence levels.

SUPPLEMENTARY DATA

Behavioral and neuronal data for error trials. We performed two sets of analysis to examine the nature of errors, and to examine cellular activity when the monkeys made errors. First, we classified errors into the following three types. The first type of error was the one in which the monkeys incorrectly selected a sequence that belonged to the same category as the correct sequence (within-category errors). The second type was defined as across- category errors when the monkeys selected a sequence belonging to other categories. The third type was the error in performing a sequence not belonging to the 11 correct sequences. In both monkeys, the occurrence of within-category errors was significantly more often than across-category errors (Chi-square test performed on all of the behavioral data taken during recordings of PF cells, $p<0.001$).

In the next analysis, we examined the extent to which neuronal activity in error trials differed from the activity during correctly performed trials. We first collected the sample of 21 cells that satisfied the following criteria. (1) Activity during preparation for one of the three categories of sequences was greater than that during preparation for others. (2) The number of either 'across-category error' or 'within category error' was greater than 3. For these cells we compared the discharge rate during the last 1 s of the preparatory period for correct trials vs. error trials. For all cells tested, the activity during preparation for 'across- category error' trials was smaller than during preparation for correct trials. The differences were statistically significant (Mann-Whitney U test performed for each of 8 cells, $p<0.01$). In contrast, for 15 cells the activity during preparation for 'within category error' trials did not differ significantly from the activity during preparation for correct trials ($p>0.05$), although for 3 cells the activity was smaller for error trials ($p<0.01$). These findings suggest that the category selective activity during the preparatory period was not observed when monkeys made an error of preparing a wrong category of sequences.

Comparison of data obtained in dorsal and ventral PF. As shown in the table below, we found that category selective cells were obtained more in the dorsal PF (Chi-square test, $p<0.01$). Cells selective for the three categories were generally more abundant in the dorsal PF. Sequence selective cells were also found more in the dorsal PF. On the other hand, statistical analysis on the strength of selectivity for the categories found no differences in the data for dorsal and ventral PF. This was confirmed by the comparison of ROC values for the paired-selective, alternate-selective, and repeat-selective cells in the dorsal vs. ventral PF (t-test, $P>0.05$).

Table 1 Category selective and sequence selective cells in dorsal and ventral PF

	Number of Recording sites	Category selective cells				Sequence selective cells
		Paired	Alternate	Repeat	Total	
Dorsal PF	120	23	24	16	63	22
Ventral PF	125	7	8	7	22	8

Category selectivity during the inter-trial interval. As explained in the description of main results, the category selectivity of PF cells started to grow during the preparatory period, culminating before the onset of the first trigger signal that prompted the first movement. A question arises as to whether the category selectivity existed to some extent before the preparatory period, during the inter-trial interval. To address this issue, we calculated the category selectivity during the epoch of the inter-trial interval for the 85 category-selective cells with the aforementioned linear regression analysis. We found that in 33 cells the category selectivity was statistically significant, although much smaller in degree in comparison to the selectivity during the preparatory period. The selectivity calculated for the 33 cells during the time epoch including the inter-trial interval is shown

in Supplementary Figure 6 (the selectivity during the period performing the preferred category of sequences and non-preferred categories are calculated separately). Thus, the category selectivity in the activity of some PF cells did not reset to zero at the end of each trial and remained during the inter-trial interval, albeit with low levels.