ARTICLES
https://doi.org/10.1038/s41593-017-0003-2
nature neuroscience

Mixed selectivity morphs population codes in prefrontal cortex

Aishwarya Parthasarathy[1,6], Roger Herikstad[2], Jit Hon Bong[2], Felipe Salvador Medina[3], Camilo Libedinsky[3,4,5*] and Shih-Cheng Yen[1,2*]

The prefrontal cortex maintains working memory information in the presence of distracting stimuli. It has long been thought that sustained activity in individual neurons or groups of neurons was responsible for maintaining information in the form of a persistent, stable code. Here we show that, upon the presentation of a distractor, information in the lateral prefrontal cortex was reorganized into a different pattern of activity to create a morphed stable code without losing information. In contrast, the code in the frontal eye fields persisted across different delay periods but exhibited substantial instability and information loss after the presentation of a distractor. We found that neurons with mixed-selective responses were necessary and sufficient for the morphing of code and that these neurons were more abundant in the lateral prefrontal cortex than the frontal eye fields. This suggests that mixed selectivity provides populations with code-morphing capability, a property that may underlie cognitive flexibility.

The prefrontal cortex plays an important role in the maintenance of working memory, as evidenced by studies using microelectrode recordings[1], lesions[2], inactivation[3], microstimulation[4,5], and functional neuroimaging[6]. Furthermore, the prefrontal cortex suppresses task-irrelevant stimuli, or distractors[3,7,8]. This has led to the notion that the prefrontal cortex plays a central role in the maintenance of working memory and the suppression of distractors, which are cornerstones of executive processing.

Individual neurons in the lateral prefrontal cortex (LPFC; Brodmann area 46) show selective, sustained activity during the delay period of working memory and only as long as the memory is maintained[1,9,10]. Populations of LPFC neurons form a stable memory code during the delay period of a working memory task[11]. Distractors are thought to be suppressed in the LPFC, since it responds less to distractors than to task-relevant stimuli[3,7,8]. Notably, sustained activity of LPFC neurons persists even after distractors are presented[3,10]. Thus, it is reasonable to hypothesize that the stable code observed during the delay period persists after a distractor is presented, forming a persistent code throughout the memory period.

Recent studies have shown that the LPFC hosts an abundance of neurons with mixed selectivity[12–19]. These cells encode multiple parameters of the task simultaneously, such as sensory stimuli, task rule, or motor response. In particular, neurons with nonlinear mixed selectivity (NMS) are thought to play a key role in the encoding of information[18,20]. In the context of a working memory task with interfering distractors, mixed selectivity could lead to a change in code after the distractor is presented. Thus, it is also reasonable to hypothesize that the code does not persist throughout the memory period but rather that it is flexible, with the ability to adapt to new task contingencies, such as the presentation of a distractor. Here we found that the LPFC morphs its code, as the latter hypothesis predicts, while the frontal eye fields (FEF) maintain a stable code, in agreement with the former hypothesis.

Results

Two monkeys were trained to perform a delayed saccade task (Fig. 1a). Overall performance of both animals was higher than 75% correct (Fig. 1b). We recorded a total of 256 neurons from the LPFC (144 from Monkey A and 112 from Monkey B; the positions of the implanted electrode arrays are shown in red in Fig. 1c) and 137 neurons from the FEF (125 from Monkey A and 12 from Monkey B; electrode arrays are shown in blue in Fig. 1c) while the animals performed the task.

Of the neurons recorded, more than 40% displayed selectivity to target location in at least one stage during the trial (Fig. 1d). Examples of the responses of an LPFC and an FEF neuron are shown in Supplementary Fig. 1a,b. To quantify the magnitude of this selectivity, we computed the percentage of explained variance (PEV) for spatial selectivity in each neuron. The average PEVs across significant neurons (Methods) are shown in Fig. 1d. In the LPFC, we observed that target information in selective cells ($n = 107$, 42% of the LPFC population) increased during the target presentation period and remained stable throughout the rest of the trial (Fig. 1d). In addition, in Supplementary Fig. 1c, we show that the distractor information was much lower in these same neurons during the Delay 2 period ($P < 0.001$, Hedges' $g = 23.99$). A previous study found a sharp decrease in target information following distractor presentation, together with an increase of distractor information[21]. Our results, however, did not replicate these observations. Rather, we found that target information remained stable, and distractor information stayed close to baseline throughout the trial (Supplementary Fig. 1c). This difference may reflect the simpler nature of our task and the comparatively lower behavioral saliency of the distractor we used. It may also reflect differences in the ways different types of information are encoded; perhaps the working memory code for numerosity in LPFC is more susceptible to distractors than the code for spatial locations. In contrast, in the FEF we observed

[1]NUS Graduate School of Integrative Science and Engineering, National University of Singapore (NUS), Singapore, Singapore. [2]Department of Electrical and Computer Engineering, NUS, Singapore, Singapore. [3]Department of Psychology, NUS, Singapore, Singapore. [4]Singapore Institute for Neurotechnology, NUS, Singapore, Singapore. [5]Institute of Molecular and Cell Biology, A*STAR, Singapore, Singapore. Present address: [6]Institute of Molecular and Cell Biology, A*STAR, Singapore, Singapore. Camilo Libedinsky and Shih-Cheng Yen contributed equally to this work.
*e-mail: camilo@nus.edu.sg; shihcheng@nus.edu.sg

that, on average, target information in selective cells ($n = 66$, 48% of the FEF population) peaked during the target presentation period ($P < 0.001$, $g = 13.27$) and dropped below the PEV values found in the LPFC for the rest of the trial ($P < 0.001$, $g = 6.54$). Similarly to the results seen for LPFC neurons, the distractor information in Delay 2 was also significantly lower than target information in these FEF neurons ($P < 0.001$, $g = 21.42$; Supplementary Fig. 1d), although for a brief period during the distractor presentation, the distractor information was not significantly different from the target information ($P \approx 0.94$). This showed that, on average, individual LPFC neurons retained more information about the target than FEF neurons. However, this analysis did not address how information about the target was stored in the population of LPFC and FEF neurons.

Distractor presentation leads to code morphing in the LPFC. We employed cross-temporal decoding methods[11] to evaluate the stability of the code used by the population to distinguish different target locations (Methods and Supplementary Fig. 2). The decoding analysis was carried out only in seven locations, since one animal had difficulty performing the task when the target was in the lower-right location. Thus, chance performance of the decoder was 1 out of 7, or 14%. In the LPFC, we found an initial period (200–650 ms) of dynamic changes in the code during the stimulus presentation and early delay period, during which the decoder performed well only when the testing windows and training windows were very close in time. Soon after, the code settled into a period of stability in Delay 1, during which the decoder performed well even when the testing windows and the training windows were separated in time (consistent with an earlier study[11]). This was maintained until the presentation of the distractor (Fig. 2a). After the distractor was presented, during Delay 2, the population settled into another period of stability, which persisted until the Go cue (Fig. 2a). Notably, the decoder performed poorly when it was trained in Delay 1 and tested in Delay 2, and vice versa (Fig. 2a), suggesting that a change occurred in the population code after the distractor presentation. To quantify this effect, we calculated the average decoding performance in the last 500 ms of each quadrant (Fig. 2a) and obtained the following decoding performance values (*LP*, LPFC performance; *FP*, FEF performance): LP_{11} (classifier trained and tested in Delay 1), LP_{12} (classifier trained in Delay 1 and tested in Delay 2), LP_{22} (classifier trained and tested in Delay 2), and LP_{21} (classifier trained in Delay 2 and tested in Delay 1). The average cross-temporal decoding performance fell from 50.5% in LP_{11} to 32.4% in LP_{12} (Supplementary Fig. 2a; $P < 0.001$, $g = 18.42$). However, LP_{22} was not significantly different from LP_{11} ($P \approx 0.50$). These results suggest that, in the LPFC, the distractor triggered a change in the Delay 1 code such that it morphed into a different code in Delay 2 with no loss in performance.

To assess whether this morphing in the code was qualitatively and quantitatively different from the temporal evolution of the code in Delay 1, it would be ideal to compare our results to an experimental condition in which the animal was required to hold the target position in working memory for the same duration but no distractor was presented. Unfortunately, this condition was not included in our experiments. In lieu of this condition, we characterized the changes in performance throughout the trial. Fig. 2b shows a heat map of the change in performance in cross-temporal decoding in the LPFC at one timepoint, t_i, and a timepoint three bins later, t_{i+3} ($LP(t_i) - LP(t_{i+3})$). For classifiers trained in Delay 1, the changes before the distractor onset were quite small (with a mean value of –0.12%), consistent with the presence of a stable population code in Delay 1. However, much larger changes (mean: 7.15%) could be seen close to the distractor offset, corresponding to a qualitative and quantitative change in the population code. For classifiers trained in Delay 2, similarly large changes (mean: –8.98%) occurred near the distractor offset (Fig. 2b); these were different from the smaller changes (mean: 1.57%) found in Delay 2. We also analyzed the rate

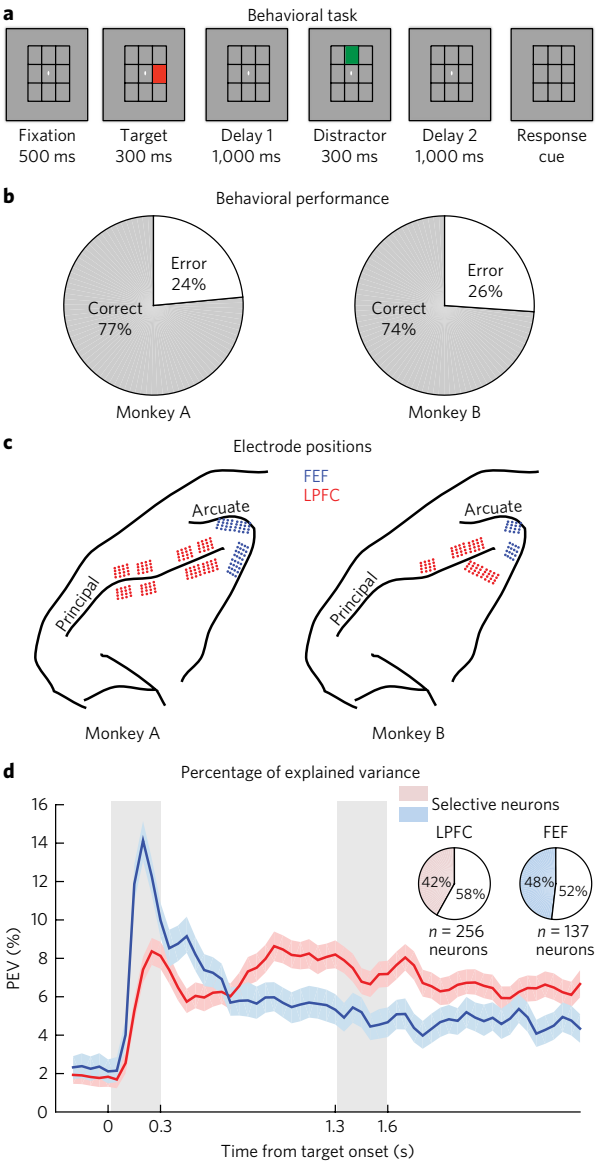

Fig. 1 | Experimental design. a, Behavioral task: each trial began when the monkey fixated on a fixation spot in the center of the screen. He was required to maintain fixation throughout the trial until the fixation spot disappeared. A target (red square) was presented for 300 ms, followed by a 1,000-ms delay period (Delay 1). A distractor (green square) was then presented for 300 ms in a random location, which was different from the target location, followed by a second delay of 1,000 ms (Delay 2). After Delay 2, the fixation spot disappeared, which was the Go cue for the monkey to report, using an eye movement, the location of the target. **b**, Pie charts indicate the performance of the two animals in our experiment. **c**, Implant locations of 16-channel and 32-channel electrode arrays (with electrode lengths ranging from 5.5 mm closer to the sulci, to 1 mm further from the sulci) in the LPFC (red dots) and the FEF (blue dots) in the two animals. **d**, Average PEV for selective neurons in the LPFC (red) and FEF (blue) plotted as a function of time in the trial. Inset: proportion of neurons in the LPFC and FEF that exhibited spatial selectivity to the target location in at least one stage during the trial. *P* values were 3.62×10^{-84} to 4.87×10^{-2} for selective LPFC neurons and 7.17×10^{-76} to 4.95×10^{-2} for selective FEF neurons. *F* values were 1.06–82.5 for selective LPFC neurons and 1.50–84.4 for selective FEF neurons. Degrees of freedom were either 6 or 7, depending on which animal the neuron was recorded from, for both LPFC and FEF. Shaded regions around the line plot indicate the 95th percentile range (two-sided). Gray shading at 0 s and 1.3 s indicates target and distractor windows, respectively.

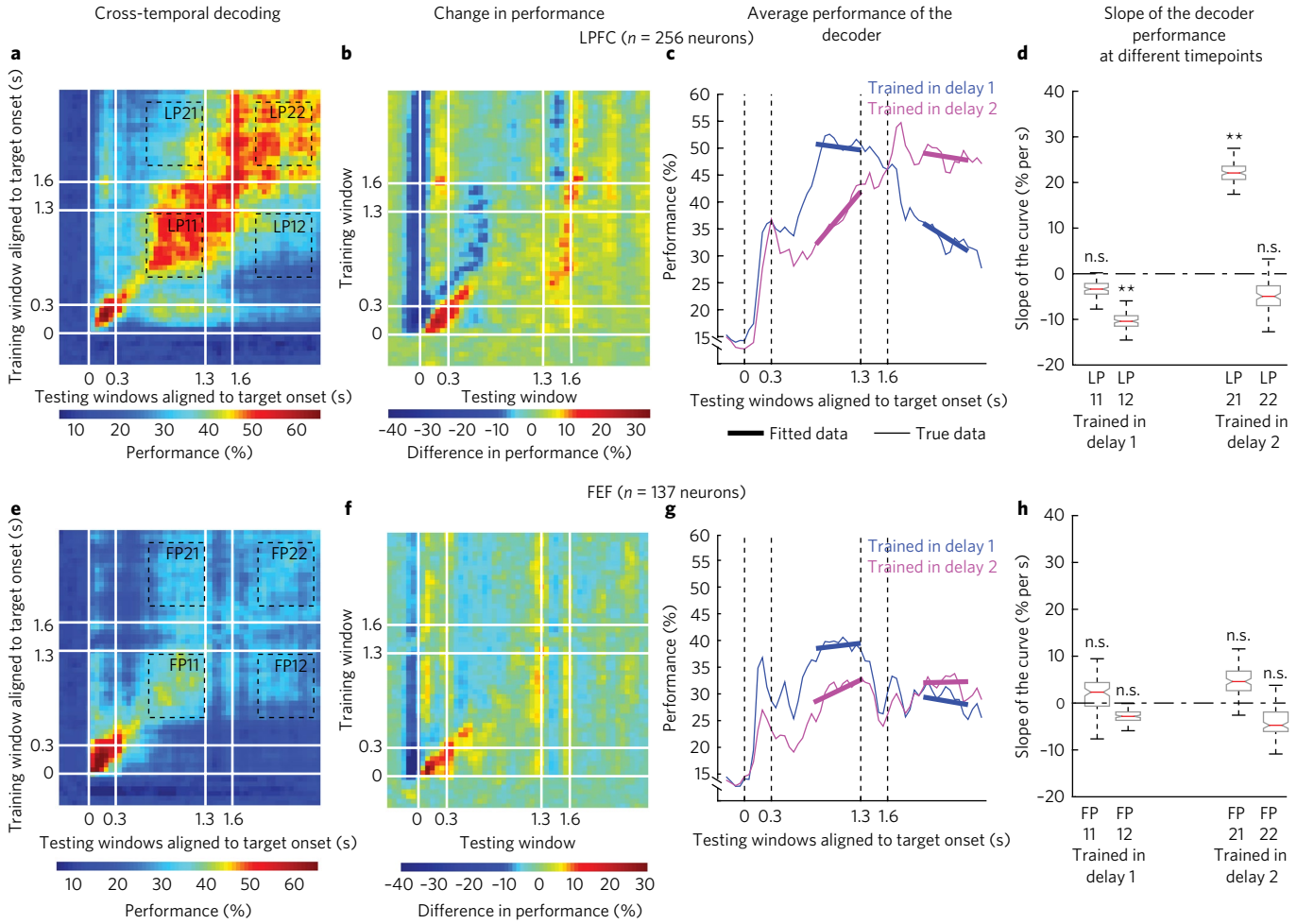

Fig. 2 | Population decoding in the LPFC and FEF. a, Heat map showing the cross-temporal population-decoding performance in the LPFC. White lines indicate target presentation (0–0.3 s) and distractor presentation (1.3–1.6 s). Squares with dashed lines indicate time periods used to quantify and compare performance results (LP_{11}, LP_{12}, LP_{22}, and LP_{21}). **b**, Heat map showing the change in performance in cross-temporal decoding in the LPFC at timepoint t_i and a timepoint three bins later, t_{i+3} ($LP(t_i) - LP(t_{i+3})$). For classifiers trained in Delay 1 and Delay 2, the changes before and after the distractor onset were quite small, while much larger changes can be seen close to the distractor offset. **c**, Average performance for LPFC classifiers trained in Delay 1 and Delay 2. The performance for Delay 1 classifiers remained fairly stable during Delay 1 but fell rapidly during Delay 2. The opposite happened for the Delay 2 classifiers. Illustrative best-fit lines (thick lines) are superimposed on the average performance curves. **d**, Distribution of the means of the slopes of the lines that were fit to the performance of the individual Delay 1 and Delay 2 classifiers when they were tested in different windows. **e**–**h**, As in (**a**–**d**) but for FEF data. In box plot: center red line, median; box limits, upper and lower quartiles; notch limits, (1.57 × interquartile range)/$\sqrt{n}$; whiskers, 95th percentile range (two sided) of the distribution. Asterisks (**), significant (i.e., 95th percentile range of the distribution did not overlap with zero); n.s., nonsignificant (i.e., 95th percentile range of the distribution overlapped with zero).

of change in performance before and after the distractor offset. Figure 2c shows the average performance for LPFC classifiers trained in Delay 1, as well as for classifiers trained in Delay 2. Best-fit lines are superimposed on the average performance curves for illustrative purposes, but in our analysis, we actually performed the fit for the individual classifiers (Supplementary Fig. 2a). Figure 2d shows the distribution of the means of the slopes of those best-fit lines in Delay 1 and Delay 2. The line fits were generally good, with the following mean R^2 values: LP_{11}, 0.24; LP_{12}, 0.74; LP_{22}, 0.2; LP_{21}, 0.92. We found that the changes in performance of the Delay 1 classifiers were very small during Delay 1 (with slope values that were not significantly different from 0, $P \approx 0.27$), but performance decreased dramatically after the distractor offset (median slope values were −10.4% per s, $P < 0.001$). Similarly, Delay 2 classifiers exhibited very small changes in performance in Delay 2 (with slope values that were not significantly different from 0, $P \approx 0.09$), but the performance decreased dramatically in Delay 1 (median slope values were 22.1% per s, $P < 0.001$). These results show that the distractor presentation clearly altered the population code in a way qualitatively and quantitatively different from the normal temporal evolution of the code.

To determine whether the morphing in population code was a general feature of prefrontal processing, we carried out the same analysis in the FEF (Fig. 2e). Similarly to our findings in the LPFC, we found an initial period (200–650 ms) of dynamic changes in the code during the stimulus presentation and early delay period, which settled into a period of stability in Delay 1. In contrast to what we found in the LPFC, after presentation of the distractor, there was a significantly smaller drop in performance from 39.0% in FP_{11} to 28.8% in FP_{12} ($P < 0.001$, $g = 12.15$). More importantly, FP_{22} was 31.7 ± 0.7%, a significant drop in performance compared to FP_{11} ($P < 0.001$, $g = 9.03$). This suggested that, unlike the LPFC, which exhibited a change in code after the distractor presentation without a loss in performance, in the FEF the code appeared to degrade in Delay 2. We also looked at the rate of change in performance of the

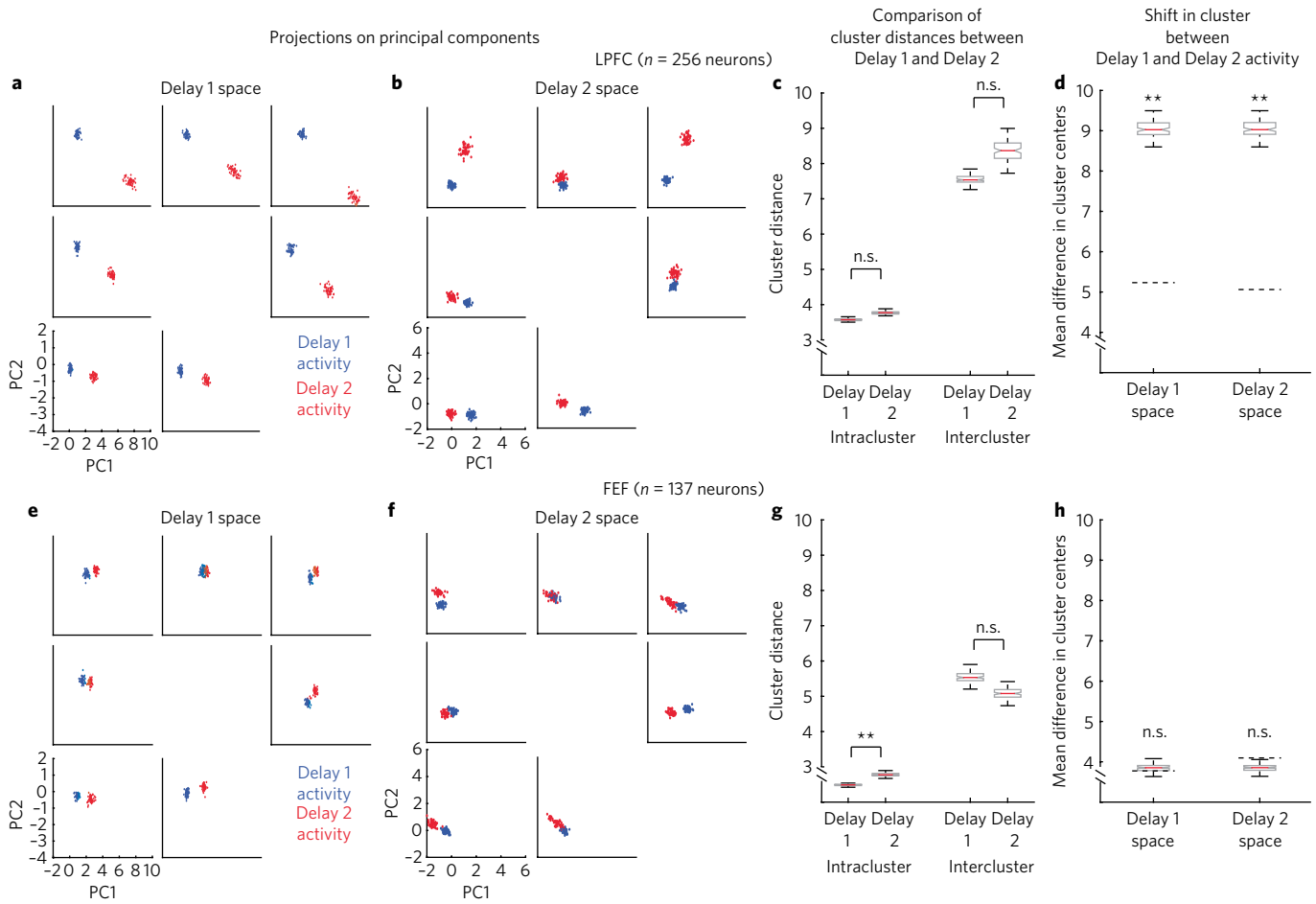

Fig. 3 | Principal components analysis (PCA) of the population code. a, Responses of the LPFC population in Delay 1 when projected onto the first two principal components (PC) of the Delay 1 response space are shown at each target location (blue). Responses of the LPFC population in Delay 2 when projected onto the same components are also shown (red). **b**, As in (**a**) except that the Delay 1 and Delay 2 responses are projected onto the first two principal components of the Delay 2 response space. **c**, Intracluster and intercluster distances (in the full PCA space) were measured for Delay 1 responses in Delay 1 space and for Delay 2 responses in Delay 2 space. There were no significant changes in either the intracluster ($P \approx 0.27$) or intercluster distances ($P \approx 0.67$), indicating that there was no loss in information after the morphing in the code. **d**, The shift in cluster centers from blue dots to red dots in (**a**,**b**) were computed for each cluster (individual cluster shifts are shown in Supplementary Fig. 2c), and the averages are plotted here for both the Delay 1 and Delay 2 spaces. The average shifts were both significantly larger than chance ($P < 0.001$; the 97.5th percentiles of the shifts obtained by chance is indicated by the horizontal dashed line; Methods). **e**–**h**, As in (**a**–**d**) but with FEF data. In box plot: center red line, median; box limits, upper and lower quartiles; notch limits, (1.57 × interquartile range)/$\sqrt{n}$; whiskers, 95th percentile range (two sided) of the distribution. Horizontal dashed lines in (**d**) and (**h**) are the 97.5th percentiles of the distribution obtained by chance. Asterisks (**), the 95th percentile ranges did not overlap in the two distributions being compared; n.s. (not significant), 95th percentile ranges overlapped in the distributions being compared.

classifiers in the FEF. Figure 2f shows the change in performance in cross-temporal decoding in the FEF at one timepoint, t_i, and a timepoint three bins later, t_{i+3} ($FP(t_i) - FP(t_{i+3})$). In contrast to what we observed in the LPFC, changes in the FEF remained fairly small, with values that remained close to 1%. Similarly, the Delay 2 classifiers exhibited very small changes during both Delay 2 (1%) and Delay 1 (2%). Figure 2g,h shows the classifier performance and the slopes of the best-fit lines for FEF classifiers. The mean R^2 values for the fits were generally good as well: FP_{11}, 0.15; FP_{12}, 0.14; FP_{22}, 0.17; FP_{21}, 0.53. However, in contrast to the LPFC, none of the slopes were significantly different from 0: FP_{11}, $P \approx 0.30$; FP_{12}, $P \approx 0.12$; FP_{22}, $P \approx 0.08$; FP_{21}, $P \approx 0.27$. These results show that, unlike in the LPFC, the population code in the FEF did not change in a qualitatively or quantitatively different way as a result of the distractor presentation.

State space analysis of code morphing. To further characterize the morphing of code seen in the LPFC, we used principal components analysis to characterize the change in the population response as a result of the presentation of the distractor (Methods). Figure 3a shows the responses in Delay 1 projected onto the first two principal components of the Delay 1 response space for each target location. For comparison, we also projected the responses in Delay 2 onto the same axes. These latter responses exhibited a clear shift from the cluster locations in Delay 1. This explains the large drop in performance when a classifier trained on Delay 1 responses was tested with Delay 2 responses. Similar shifts were found for Delay 2 responses and Delay 1 responses when projected onto the top two principal components of the Delay 2 response space (Fig. 3b). To relate the cluster plots to the performance we observed in the cross-temporal decoding in Figure 2a, we computed the average distance between responses in the full principal component analysis space for the same target location (intracluster distance) and the average distance between responses for different target locations (intercluster distance). The first measure quantifies the intracondition response

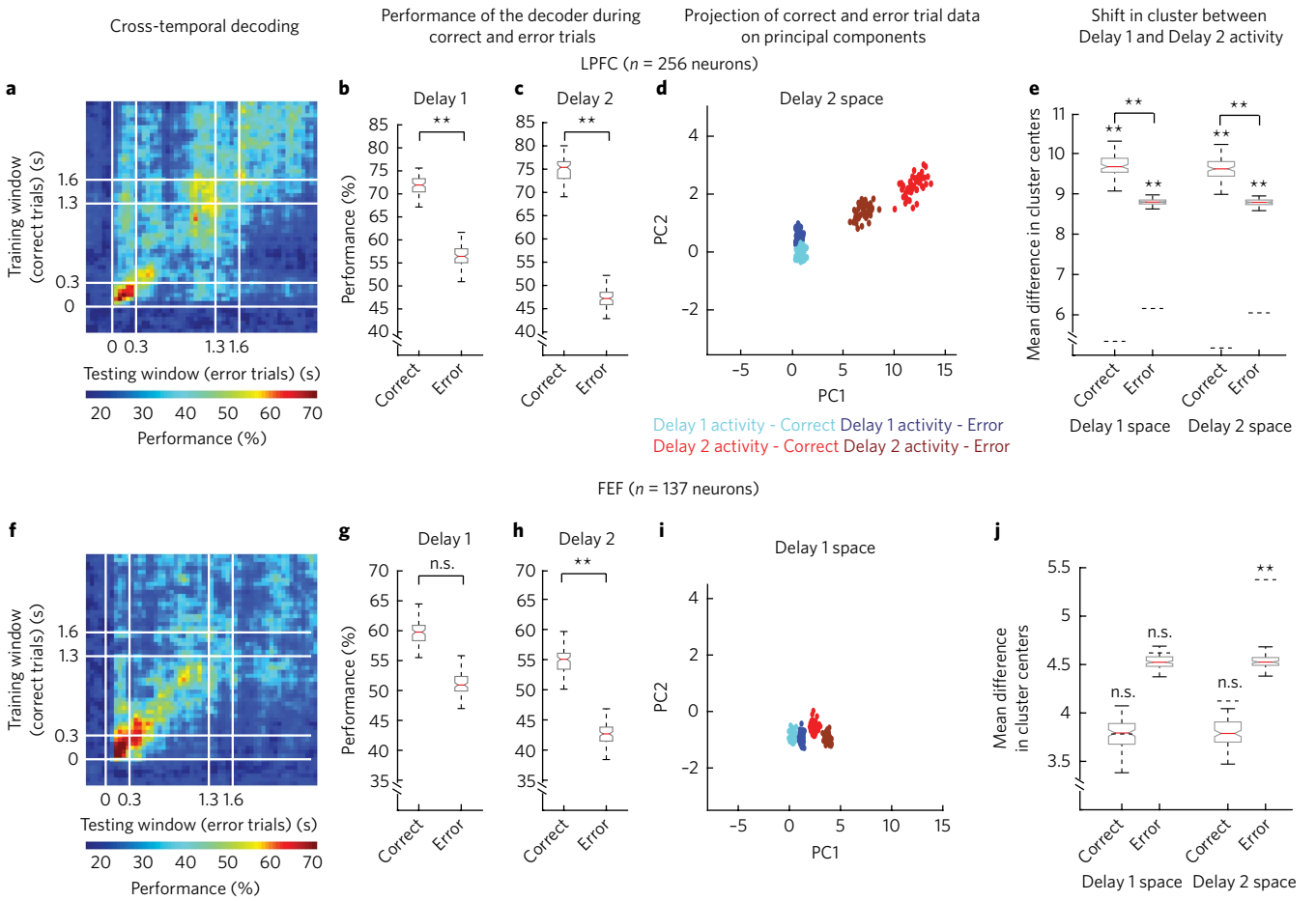

Fig. 4 | Decoding of error trials. a, Heat map showing the cross-temporal decoding performance (for four target locations) in the LPFC for error trials. **b**, Performance in a 500-ms window in Delay 1 (LP_{11}) for error trials was significantly lower than for correct trials ($P < 0.001$). **c**, Performance in a 500-ms window in Delay 2 (LP_{22}) for error trials was significantly lower than for correct trials ($P < 0.001$). **d**, Responses of the LPFC population in Delay 2 when projected onto the first two principal components of the Delay 2 responses (correct and error trials: red and brown, respectively). Projections for one target location are shown. Responses of the LPFC population in Delay 2 for the same location when projected onto the same components are also shown (cyan and blue for correct and error trials, respectively). **e**, The shift in the cluster centers from Delay 1 to Delay 2 in Delay 1 space (as in **d**) were computed for each cluster, and the averages are plotted here. The shifts in Delay 2 space are also shown. Horizontal dashed line indicates the 97.5th percentile of the shifts obtained by chance. **f**–**j**, As in (**a**–**e**) but with FEF data. In box plot: center red line, median; box limits, upper and lower quartiles; notch limits, (1.57 × interquartile range)/$\sqrt{n}$; whiskers, 95th percentile range (two sided) of the distribution. Horizontal dashed lines in (**e**) and (**j**) are the 97.5th percentile ranges of the distributions obtained by chance. Asterisks (**), the 95th percentile ranges (two-sided) did not overlap in the two distributions being compared; n.s. (not significant), 95th percentile ranges overlapped in the distributions being compared.

variability found at each target location, while the second quantifies the intercondition discriminability between target locations. The results are plotted in Figure 3c and show that the intracluster and intercluster distances for Delay 1 responses in Delay 1 space and for Delay 2 responses in Delay 2 space were not significantly different ($P \approx 0.67$ for intercluster distance, $P \approx 0.27$ for intracluster distance). This provides additional evidence that there was no loss in information after the code morphing. We also quantified the shift in cluster locations between the Delay 1 and Delay 2 responses illustrated in Figure 3a,b and found shifts that were significantly larger than chance in Figure 3d (Delay 1 space: $P < 0.001$, $g = 24.72$; Delay 2 space: $P < 0.001$, $g = 24.89$). These results suggest that the morphed code in the population underwent a transformation that maintained the intracondition variability and intercondition discriminability, thus allowing the LPFC to maintain information after the presentation of the distractor.

In contrast, we found that in the FEF, unlike in the LPFC, there were much smaller shifts in the cluster positions (Fig. 3e,f). There were no significant increases in intercluster distances ($P \approx 0.91$), although there was a significant increase in intracluster distance ($P < 0.001$, $g = 1.91$; Fig. 3g). In addition, we found the shifts in cluster positions to be no larger than chance ($P \approx 0.43$ for Delay 1 and $P \approx 0.99$ for Delay 2; Fig. 3h). This meant that the code morphing in the FEF was very small, and the overall shift was not significant. Coupled with a significant increase in intracluster distances, it appeared that the population code in Delay 2 was a degraded version of the population code in Delay 1. After the Go cue, target information reappeared in the FEF before saccade onset, underscoring the role of the FEF in controlling eye movements (Supplementary Fig. 3).

Overall, these results suggest that the presentation of the distractor triggered reorganization of the target information in both the LPFC and FEF, but the LPFC was able to morph its population code in a way that retained information, while the population code in the FEF appeared to persist across delay periods, although with substantial loss of information. We also observed these shifts (as well as

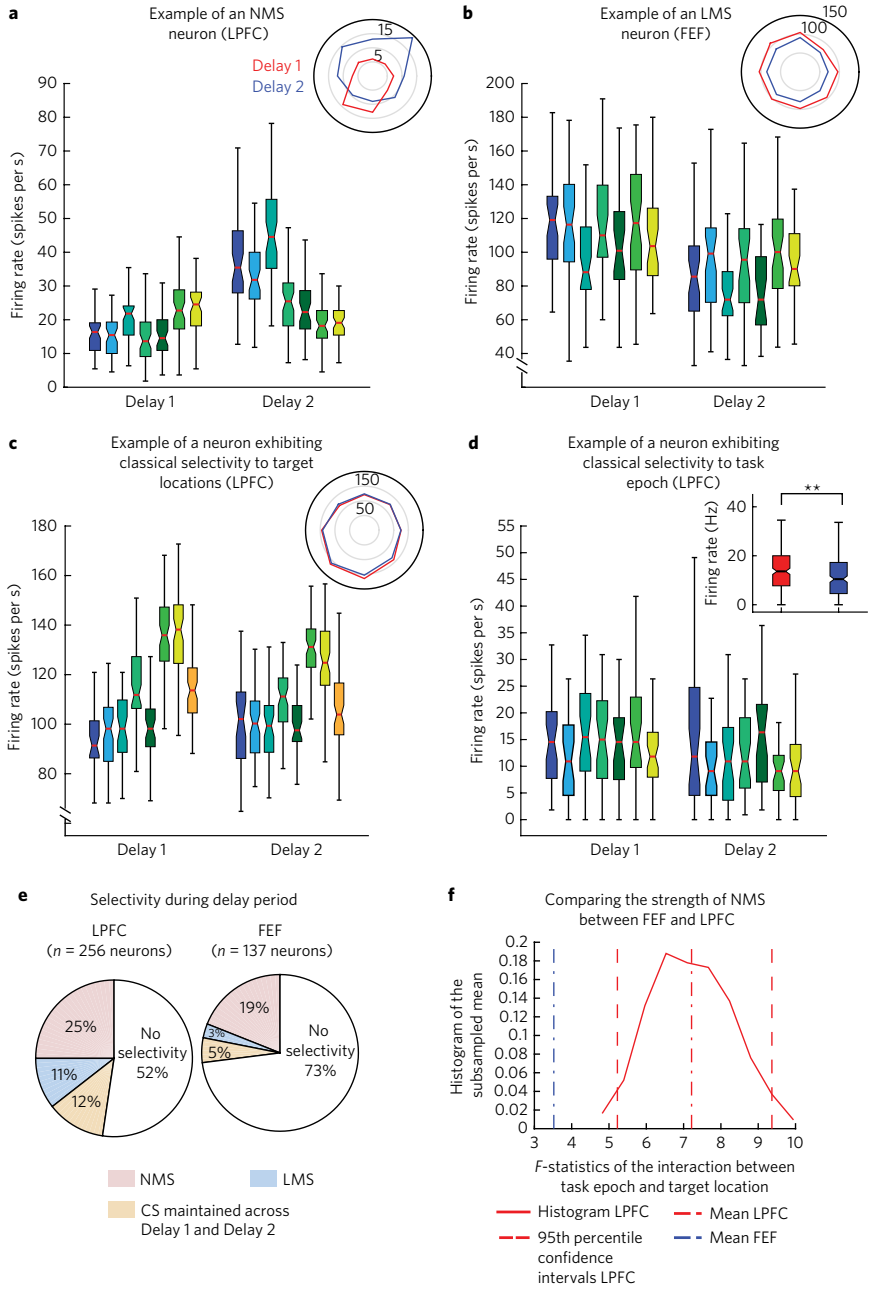

Fig. 5 | Mixed selective responses. a, An example of a neuron with NMS is shown. Box plots illustrate the firing rates at different target locations (represented by different colors) in Delays 1 and 2; polar plots (inset) show the spatial tuning curves in Delay 1 (red) and Delay 2 (blue). Borders of the box plots indicate the 25th and 75th percentiles of the firing rate distribution; whiskers indicate the 95th percentile range (two-sided) of the distributions. Horizontal red lines indicate the medians of the distribution. This neuron showed a fairly large change in spatial selectivity, as well as an overall increase in mean firing rate ($P=1.234\times10^{-17}$, $P=3.864\times10^{-6}$, and $P=1.387\times10^{-33}$ for main effect of target location, main effect of task epoch, and the interaction term, respectively). **b**, An example of a neuron with LMS. This neuron exhibited an overall decrease in mean firing rate, but its selectivity did not change ($P=0.0314$, $P=0.0068$, and $P=0.4036$ for main effect of target location, main effect of task epoch, and the interaction term, respectively). **c**, An example of a neuron with CS to target location ($P=0.0386$, $P=0.4902$, and $P=0.9526$ for main effect of target location, main effect of task epoch, and the interaction term, respectively). **d**, Example of a neuron with CS to task epoch ($P=0.8767$, $P=0.0034$, and $P=0.6060$ for main effect of target location, main effect of task epoch, and the interaction term, respectively). Inset: firing rates averaged across target locations in Delay 1 (red) and Delay 2 (blue). **e**, Distributions of neurons with different response selectivities in the LPFC and FEF. Neurons with NMS in the LPFC exhibited P values for the interaction between 4.54×10^{-47} and 4.97×10^{-2}, while those in the FEF exhibited P values between 6.55×10^{-6} and 4.41×10^{-2}. Neurons with LMS in the LPFC exhibited P values between 9.09×10^{-31} and 3.98×10^{-2} for target location and between 1.37×10^{-223} and 2.50×10^{-2} for task epoch; those in the FEF exhibited P values between 1.03×10^{-20} and 4.33×10^{-2} for target location and between 6.06×10^{-29} and 1.52×10^{-2} for task epoch. Neurons with CS in the LPFC exhibited P values between 1.43×10^{-15} and 4.20×10^{-2} for target location or task epoch; those in the FEF exhibited P values between 3.88×10^{-18} and 4.43×10^{-2} for target location or task epoch. LPFC and FEF neurons exhibited either 6 or 7 degrees of freedom for the interaction term depending on the animal from which the neuron was recorded. **f**, Using the F statistic, we also compared the effect sizes found in the neurons with NMS in the LPFC and FEF. The distribution of effect sizes found in the LPFC is shown by the solid red line, with the mean and the 95th percentile range (two-sided) shown with the red dashed vertical lines. The mean of the FEF population is shown by the blue dashed vertical line, which is clearly outside the 95th percentile for the LPFC.

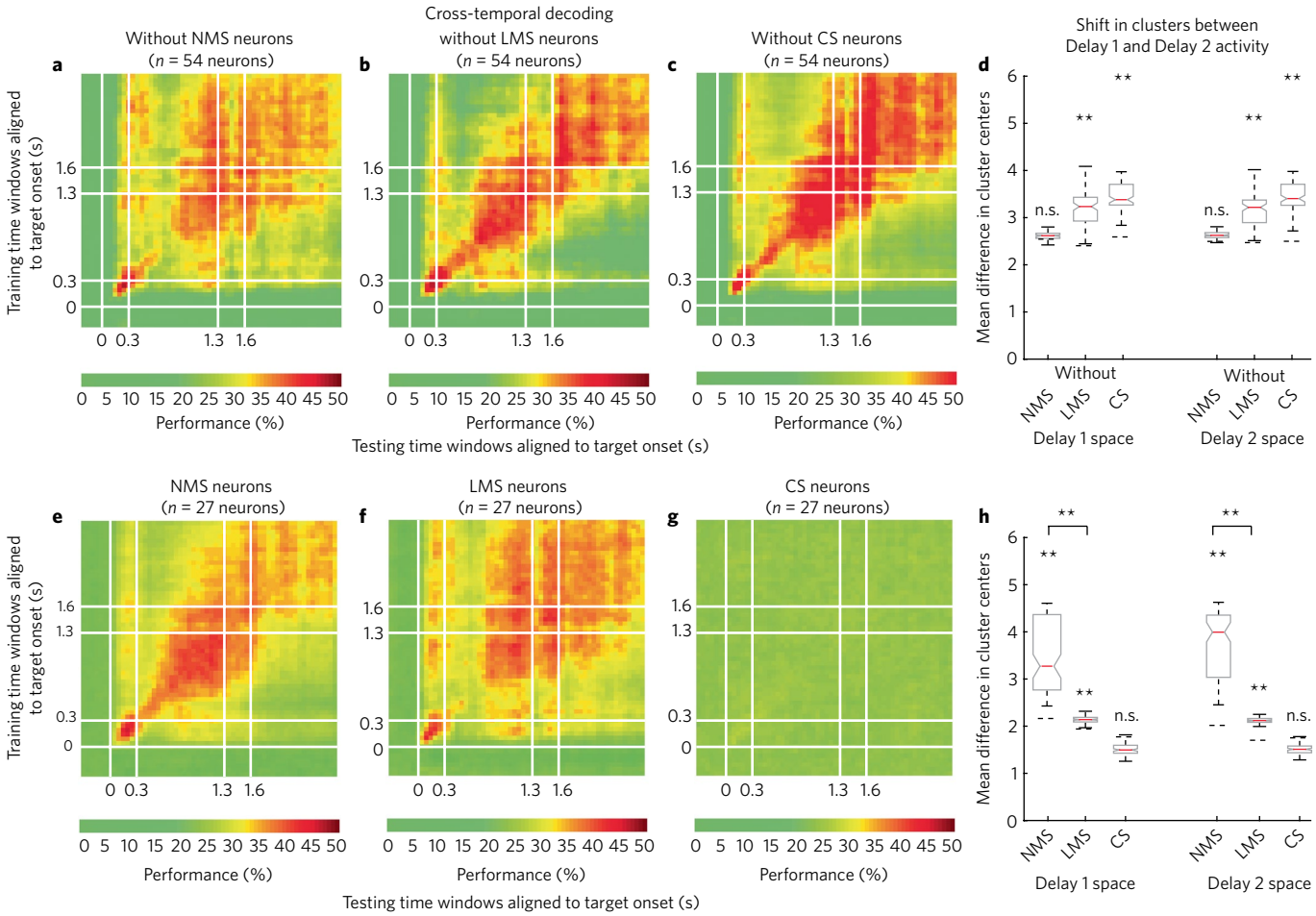

Fig. 6 | Roles of different types of neurons in code morphing. We constructed a pseudopopulation of LPFC neurons, consisting of equal numbers of neurons with CS, LMS, and NMS (27 neurons each). **a**, Cross-temporal decoding performance of a population without NMS neurons. Code morphing was eliminated, as shown by the lack of two distinct stable codes. **b**,**c**, Cross-temporal decoding performance of a population without LMS neurons and one without CS neurons. In both cases, the code morphing was preserved. **d**, The shift in the cluster centers across delays for pseudopopulations without NMS, without LMS, or without CS. Without NMS, the shift was not significant (Delay 1: $P \approx 0.53$; Delay 2: $P \approx 0.40$; Supplementary Fig. 10a,b). Without LMS, the shift was significant for both Delay 1 ($P < 0.001$, $g = 3.06$) and Delay 2 ($P < 0.001$, $g = 3.25$; Supplementary Fig. 10e,f). Without CS, the shift was significant for both Delay 1 ($P < 0.001$, $g = 6.32$) and Delay 2 ($P < 0.001$, $g = 6.16$; Supplementary Fig. 10i,j). **e**, Cross-temporal decoding performance of a population consisting only of NMS neurons. The code morphing was largely intact. **f**,**g**, Cross-temporal decoding performances of a population consisting only of LMS neurons and one consisting only of CS neurons. In both cases, the code morphing was greatly reduced or entirely absent. **h**, The shift in the cluster centers across delays for a pseudopopulation comprised only of NMS, LMS, or CS neurons. Neurons with NMS were responsible for a large change in code (Delay 1: $P < 0.001$, $g = 5.93$; Delay 2: $P < 0.001$, $g = 7.17$; Supplementary Fig. 10c,d), neurons with LMS exhibited a small change in code (Delay 1: $P < 0.001$, $g = 1.23$; Delay 2: $P < 0.001$, $g = 2.41$; Supplementary Fig. 10g,h), and neurons with CS had no role in changing the code (Delay 1: $P \approx 0.98$; Delay 2: $P \approx 0.99$; Supplementary Fig. 10k,l). The change in code in LMS neurons was significantly lower than that in NMS neurons (Delay 1: $P < 0.001$, $g = 1.31$; Delay 2: $P < 0.001$, $g = 2.74$). In box plot: center red line, median; box limits, upper and lower quartiles; notch limits, $(1.57 \times \text{interquartile range}) / \sqrt{n}$; whiskers, 95th percentile range (two sided) of the distribution. Horizontal dashed lines in the box plot represent the 97.5th percentile of the shifts obtained by chance. Asterisks (**), the 95th percentile ranges (two-sided) did not overlap between in the two distributions being compared; n.s. (not significant), 95th percentile ranges overlapped in the distributions being compared.

the cross-temporal decoding results described above) when using simultaneously recorded populations of neurons instead of pseudopopulations (Supplementary Fig. 4). In addition, the code morphing was also observed using a cross-correlation analysis of target PEVs[22] (Supplementary Fig. 5). We did not find any evidence that the code morphing in the LPFC was dependent on the distractor location (Supplementary Figs. 6 and 7), although our analysis was underpowered, as we did not have enough repetitions to cover the large number of target–distractor combinations (Methods). Finally, we found that the LPFC encoded low but stable levels of distractor information in Delay 2 (Supplementary Fig. 1e), while the FEF encoded distractor information only during the distractor presentation period (Supplementary Fig. 1f).

Temporal stability in error trials. To assess the behavioral relevance of the code morphing observed in the LPFC, we analyzed trials in which the animals waited for the Go cue but failed to report the correct target location (error trials). Due to the small number of error trials at some locations, we tested the decoding on just four of the locations with the highest rate of errors (Methods). We decoded the location of the target in error trials (the distribution of different types of error trials is shown in Supplementary Fig. 8).

For each time window t_i in the error trials, we used the classifier trained in the same time window on correct trials and tested it across all time windows in error trials. Figure 4a shows the results of the cross-temporal decoding of error trials in the LPFC. Notably, the difference in performance of correct and incorrect trials in the initial stages of the response ($t = 200$–650 ms) was not significant (correct trials = 52.4 ± 0.2%; error trials = 48.9 ± 0.1%; $P \approx 0.06$), perhaps due to the fact that the initial response was dominated by the visual response (as seen in the similarity between 200 ms and 650 ms in Figs. 2a and 4a). However, the mean performance of the decoder during Delay 1 (LP_{11}) in error trials was significantly lower than in correct trials ($P < 0.001$, $g = 7.37$; Fig. 4b). This suggested that in error trials, information in Delay 1 about the target location had already significantly deteriorated. LP_{22} was also significantly lower in error trials compared to correct trials ($P < 0.001$, $g = 11.29$; Fig. 4c), but the difference between LP_{11} and LP_{22} in error trials was not significant ($P \approx 0.12$, data not shown). In Figure 4d, we show a comparison of code morphing in error trials to that in correct trials for an example target location (error trials for other target locations are shown in Supplementary Fig. 9). In some cases (e.g., Supplementary Fig. 9a; top middle location), the error trials came very close to the correct trials, while in others (e.g., bottom left location), they were more widely separated. In addition, the responses in Delay 2 were also projected onto the same space, with correct trials plotted using red dots and error trials plotted using brown dots. The plots show that, even in error trials, there were shifts in the cluster locations for many of the target locations, indicating code morphing. Notably, in some cases, the clusters in the error trials were highly overlapped with the clusters in correct trials (e.g., top middle and bottom left locations), while in others there was a clear separation (e.g., top right corner and middle right location), which suggested an incomplete morphing of the code in error trials. We quantified the average shift in the cluster centers between the Delay 1 and Delay 2 responses across the four target locations, and the shifts in error trials were significantly smaller than in correct trials in both the Delay 1 and Delay 2 spaces (Fig. 4e; Delay 1: $P < 0.001$, $g = 24.89$ for correct trials when compared to chance; $P < 0.001$, $g = 20.69$ for error trials when compared to chance; $P < 0.001$, $g = 6.52$ when comparing correct trials to error trials; Delay 2: $P < 0.001$, $g = 24.73$ for correct trials when compared to chance; $P < 0.001$, $g = 19.57$ for error trials when compared to chance; $P < 0.001$, $g = 6.86$ when comparing correct trials to error trials). This suggests that incomplete code morphing may have been the cause of some of the errors.

In comparison, in the FEF we found no significant differences in the cross-temporal decoding performance for correct and error trials in the target period (correct trials = 63.5 ± 0.2%, error trials = 60.6 ± 0.2%, $P \approx 0.40$; Fig. 4f) and Delay 1 period ($P \approx 0.06$; Fig. 4g), although the differences during the Delay 2 period were significant ($P < 0.001$, $g = 6.01$; Fig. 4h). Notably, the decrease in decoding performance during Delay 2 in the FEF was significantly smaller than that observed in the LPFC (FEF: difference between correct and error = 12.2 ± 0.3%; LPFC: difference between correct and error = 27.7 ± 0.3%, $P < 0.001$, $g = 4.78$). This suggests that activity in the LPFC during the delay period was more closely related to the animal's behavior than activity in the FEF. In Fig. 4i,j, we show the lack of code morphing in the FEF in error trials, much like what we found in correct trials (Delay 1: $P \approx 0.53$ for correct trials; $P \approx 0.72$ for error trials; Delay 2: $P \approx 0.74$ for correct trials when compared to chance; $P < 0.001$, $g = 14.38$ for error trials when compared to chance, which means that there was significantly less morphing than expected by chance). Overall, our results show that, similarly to the situation in correct trials, code morphing occurred in the LPFC but not in the FEF in error trials. However, subtle differences in the population responses in the LPFC were enough to cause errors to occur.

Neurons with mixed selective activity morph the population code. One possible mechanism that could explain the morphing in code in the LPFC involves single-neuron changes in responsiveness and selectivity in different task epochs, referred to as mixed-selective responses[18,23]. We performed two-way ANOVA on the neuronal firing rates between task epoch (Delay 1 and 2) × target location (seven locations) and divided the neurons into the following three categories: (i) neurons with classical selectivity (CS; i.e., no mixed selectivity), which exhibited a main effect of either target location or task epoch; (ii) neurons with linear mixed-selectivity (LMS), which exhibited main effects of both target location and task epoch but a nonsignificant interaction term; and (iii) neurons with nonlinear mixed selectivity (NMS), which exhibited a significant interaction term. Figure 5a–d shows examples of each type of neuron. In total, 25% of LPFC neurons ($n = 64$) and 19% of FEF neurons ($n = 26$) were identified as neurons with NMS; 11% of LPFC neurons ($n = 27$) and 3% of FEF neurons ($n = 18$) were identified as neurons with LMS; and 12% of LPFC neurons ($n = 30$) and 5% of FEF neurons ($n = 6$) were identified as neurons with CS (Fig. 5e). In addition, the strength of the interaction (as measured by F-statistics from the two-way ANOVA; Methods) in these neurons with NMS was significantly higher in the LPFC than the FEF ($P < 0.001$; Fig. 5f). This meant not only that there were more neurons with NMS in the LPFC than in the FEF but also that the LPFC neurons exhibited larger changes in their selectivity between Delays 1 and 2. This difference suggested that the morphing in code observed in the LPFC could be driven by neurons with mixed selectivity.

We tested this hypothesis by assessing the roles of these three types of neurons in mediating code morphing in the LPFC. We did this by systematically eliminating each type of neuron from the LPFC population (Methods) and performed cross-temporal decoding on the remaining population. When we eliminated the neurons with NMS, code morphing was no longer observable (Fig. 6a,d). On the other hand, when we eliminated the neurons with LMS or CS, code morphing was unaffected (Fig. 6b–d). When we carried out the same analysis separately on homogeneous populations of neurons with NMS, LMS, and CS, we found that neurons with NMS were responsible for a large change in code, neurons with LMS exhibited a small change in code, and neurons with CS had no role in changing the code (Fig. 6e–h and Supplementary Fig. 10). However, the code change in LMS neurons was significantly lower than that in NMS neurons (Fig. 6h). To assess the effect that neurons with NMS would have on the FEF, we added LPFC neurons with NMS to the FEF population to match the proportion and properties found in the LPFC, and we found that the code morphing could now be observed in the FEF (Supplementary Fig. 11a). In addition, we removed the most nonlinear neurons with NMS from the LPFC until the population of neurons with NMS matched the properties of those found in the FEF and found code-morphing results similar to those found in the FEF (Supplementary Fig. 11b). These results provided strong evidence that neurons with mixed selectivity (NMS and LMS) were necessary and sufficient for the code morphing in the LPFC.

Our definition of neurons with NMS included neurons that changed selectivity between Delay 1 and Delay 2 (i.e., Fig. 5a), as well as neurons that were selective in Delay 1 but lost their selectivity in Delay 2, and vice versa. Excluding neurons that acquired or lost selectivity from the analysis did not substantially change the results (Supplementary Fig. 12). This shows that the morphing of the population code was not exclusively driven by neurons that acquired or lost selectivity.

In addition to the population of neurons with NMS that encoded a mixture of target location and delay period, we identified another population of neurons with NMS that encoded a mixture of target location and distractor location (Supplementary Fig. 13). This population, however, was not involved in the morphing of code observed in the LPFC (Supplementary Fig. 14). Finally, we also compared the

receptive field sizes in the LPFC and FEF (Supplementary Fig. 15) and were not able to find any significant differences in size ($P \approx 0.6$ and $P \approx 0.54$, respectively, for Delays 1 and 2) that could explain the differences in code morphing in the LPFC compared to the FEF.

Discussion

Here we demonstrate that the LPFC encodes memory information in a flexible manner, by morphing its memory code in response to a distractor, without losing information. The morphing of code is behaviorally relevant, since incomplete morphing was observed in error trials. We identified the cellular mechanism that underlies the morphing of the code by determining that neurons with mixed selectivity are necessary and sufficient for the morphing to occur. In particular, neurons with NMS had the largest influence in the morphing of the code. Finally, we show that unlike the LPFC, the FEF, another prefrontal region, encodes memory information using a consistent code before and after the distractor (i.e., without morphing its code) but that the FEF code was unstable, since it lost information after the distractor was presented. Overall, our results show that the balance between classical selectivity and mixed selectivity within a neuronal population determines whether the population encodes information consistently throughout the task, or flexibly, by morphing the code in response to task contingencies, such as distractors.

Previous studies have shown that cells with NMS increase the dimensionality of neuronal representations, thus optimizing the encoding of information[24–26]. This increased dimensionality allows a neuronal population to encode information in a manner that could be flexibly decoded, or interpreted, by downstream neurons[18,21]. Our results provide a previously uncharacterized role for NMS: to allow the morphing of codes in a manner that maintains the encoded information, even when the morphing is triggered by a task-irrelevant stimulus, or distractor. Thus, in addition to the proposed roles of increasing dimensionality and multiple-interpretability of the code, we propose that NMS allows neuronal populations to flexibly adjust the code itself, a role that fits well with the adaptive coding model of neural function in prefrontal cortex[27] and with dynamic network models of prefrontal function[23,28]. Altogether, a picture is emerging of flexibility in prefrontal coding, which may well be a neural mechanism for cognitive flexibility.

It has also been shown that the LPFC plays a different role than visual and parietal areas do in working memory and attention[3,10,21,29–33]. Neurons in both the LPFC and FEF show sustained activity during the delay period of a working memory task. However, the differences in the roles of these two prefrontal regions in working memory have, thus far, not been explored. Our results show clear differences between the LPFC and FEF. First, only the LPFC morphed its code after a distractor. Second, only the LPFC maintained the same amount of information after the distractor was presented. Third, the LPFC showed a larger decrease in memory information during error trials. Finally, while neurons with NMS were present in both the LPFC and FEF, they were less abundant and weaker in the FEF than in the LPFC, which would explain why we observed a consistent code in the FEF before and after the distractor. Overall, our results underscore the large functional differences between the LPFC and FEF, with LPFC activity being more closely related to the maintenance of working memory.

Although there were clear signs that the change in code was triggered by the presentation of the distractor, it is a limitation in our experimental design that we did not have a condition without a distractor. Although we believe our results strongly suggest that the code morphing was due to the presentation of the distractor (especially since previous studies without a distractor[22,34] have demonstrated the existence of a persistent, stable code beyond 1-s delays), this remains to be tested explicitly with future experiments without distractors.

Our results also raise a number of questions. For example, would different types of distractors (e.g., auditory) morph the code in the LPFC differently? Would subsequent distractors continue morphing the code even further? Are there any links between the multiple network states that encode the same memories in different contexts? Is the morphing in code gradual or abrupt in individual trials? How are areas downstream from the LPFC able to interpret the changing code correctly? Future studies should address these and other questions raised by this work.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41593-017-0003-2.

Received: 9 January 2017; Accepted: 1 September 2017;
Published online: 9 October 2017

References

1. Fuster, J. M. & Alexander, G. E. Neuron activity related to short-term memory. *Science* **173**, 652–654 (1971).
2. Funahashi, S. Functions of delay-period activity in the prefrontal cortex and mnemonic scotomas revisited. *Front. Syst. Neurosci* **9**, 2 (2015).
3. Suzuki, M. & Gottlieb, J. Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. *Nat. Neurosci.* **16**, 98–104 (2013).
4. Stamm, J. S. Electrical stimulation of monkeys' prefrontal cortex during delayed-response performance. *J. Comp. Physiol. Psychol.* **67**, 535–546 (1969).
5. Wegener, S. P., Johnston, K. & Everling, S. Microstimulation of monkey dorsolateral prefrontal cortex impairs antisaccade performance. *Exp. Brain Res.* **190**, 463–473 (2008).
6. Cohen, J. D. et al. Activation of the prefrontal cortex in a nonspatial working memory task with functional MRI. *Hum. Brain Mapp.* **1**, 293–304 (1994).
7. Everling, S., Tinsley, C. J., Gaffan, D. & Duncan, J. Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nat. Neurosci.* **5**, 671–676 (2002).
8. Sakai, K., Rowe, J. B. & Passingham, R. E. Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nat. Neurosci.* **5**, 479–484 (2002).
9. Romo, R., Brody, C. D., Hernández, A. & Lemus, L. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* **399**, 470–473 (1999).
10. Miller, E. K., Erickson, C. A. & Desimone, R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* **16**, 5154–5167 (1996).
11. Stokes, M. G. et al. Dynamic coding for cognitive control in prefrontal cortex. *Neuron* **78**, 364–375 (2013).
12. Asaad, W. F., Rainer, G. & Miller, E. K. Neural activity in the primate prefrontal cortex during associative learning. *Neuron* **21**, 1399–1407 (1998).
13. Mansouri, F. A., Matsumoto, K. & Tanaka, K. Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. *J. Neurosci.* **26**, 2745–2756 (2006).
14. Hussar, C. R. & Pasternak, T. Flexibility of sensory representations in prefrontal cortex depends on cell type. *Neuron* **64**, 730–743 (2009).
15. Warden, M. R. & Miller, E. K. Task-dependent changes in short-term memory in the prefrontal cortex. *J. Neurosci.* **30**, 15801–15810 (2010).
16. Noudoost, B. & Moore, T. The role of neuromodulators in selective attention. *Trends Cogn. Sci.* **15**, 585–591 (2011).
17. Liebe, S., Hoerzer, G. M., Logothetis, N. K. & Rainer, G. Theta coupling between V4 and prefrontal cortex predicts visual short-term memory performance. *Nat. Neurosci.* **15**, 456–462 (2012). S1–S2.
18. Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
19. Sarma, A., Masse, N. Y., Wang, X.-J. & Freedman, D. J. Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. *Nat. Neurosci.* **19**, 143–149 (2016).
20. Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* **37**, 66–74 (2016).
21. Jacob, S. N. & Nieder, A. Complementary roles for primate frontal and parietal cortex in guarding working memory from distractor stimuli. *Neuron* **83**, 226–237 (2014).
22. Murray, J. D. et al. Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci. USA* **114**, 394–399 (2017).

23. Enel, P., Procyk, E., Quilodran, R. & Dominey, P. F. Reservoir computing properties of neural dynamics in prefrontal cortex. *PLOS Comput. Biol.* **12**, e1004967 (2016).
24. Miller, E. K. & Fusi, S. Limber neurons for a nimble mind. *Neuron* **78**, 211–213 (2013).
25. Sreenivasan, K. K., Curtis, C. E. & D'Esposito, M. Revisiting the role of persistent neural activity during working memory. *Trends Cogn. Sci.* **18**, 82–89 (2014).
26. Stokes, M. G. 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* **19**, 394–405 (2015).
27. Duncan, J. An adaptive coding model of neural function in prefrontal cortex. *Nat. Rev. Neurosci.* **2**, 820–829 (2001).
28. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
29. Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K. & Poggio, T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J. Neurophysiol.* **100**, 1407–1419 (2008).
30. Buschman, T. J. & Miller, E. K. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* **315**, 1860–1862 (2007).
31. Qi, X.-L. et al. Comparison of neural activity related to working memory in primate dorsolateral prefrontal and posterior parietal cortex. *Front. Syst. Neurosci* **4**, 12 (2010).
32. Salazar, R. F., Dotson, N. M., Bressler, S. L. & Gray, C. M. Content-specific fronto-parietal synchronization during visual working memory. *Science* **338**, 1097–1100 (2012).
33. Qi, X. L., Elworthy, A. C., Lambert, B. C. & Constantinidis, C. Representation of remembered stimuli and task information in the monkey dorsolateral prefrontal and posterior parietal cortex. *J. Neurophysiol.* **113**, 44–57 (2015).
34. Spaak, E., Watanabe, K., Funahashi, S. & Stokes, M. G. Stable and dynamic coding for working memory in primate prefrontal cortex. *J. Neurosci.* **37**, 6503–6516 (2017).

Acknowledgements

We thank A. Tan for comments and suggestions on an earlier version of this manuscript. We thank C. Lim, M.N. Lynn, L. Chan, K. Chng, and E.M. Peña for help with animal training, surgery, and care. This work was supported by startup grants from the Ministry of Education Tier 1 Academic Research Fund and SINAPSE to C.L., a grant from the NUS-NUHS Memory Networks Program to S.-C.Y., and a grant from the Ministry of Education Tier 2 Academic Research Fund to C.L. and S.-C.Y. (MOE2016-T2-2-117).

Author contributions

C.L., A.P., R.H., J.H.B., and S.-C.Y. designed the experiments. A.P., R.H., and J.H.B. collected behavioral and electrophysiological data. F.S.M. and A.P. performed the microstimulation verification experiments. A.P. and R.H. analyzed behavioral and electrophysiological data. S.-C.Y. and C.L. guided the data analysis. All authors discussed the results, and A.P., C.L., and S.-C.Y. wrote the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-017-0003-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.L. or S.-C.Y.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Subjects and surgical procedures. We used two male adult macaques (*Macaca fascicularis*) in the experiments: Animal A (age 4) and Animal B (age 6). All animal procedures were approved by, and conducted in compliance with, the standards of the Agri-Food and Veterinary Authority of Singapore and the Singapore Health Services Institutional Animal Care and Use Committee (SingHealth IACUC #2012/SHS/757). Procedures also conformed to the recommendations described in *Guidelines for the Care and Use of Mammals in Neuroscience and Behavioral Research* (National Academies Press, 2003). Each animal was implanted first with a titanium head-post (Crist Instruments, MD, USA) before arrays of intracortical microelectrodes (MicroProbes, MD, USA) were implanted in multiple regions of the left frontal cortex (Fig. 1c). In Animal A, we implanted six arrays of 16 electrodes and one array of 32 electrodes in the LPFC, as well as two arrays of 32 electrodes in the FEF, for a total of 192 electrodes. In Animal B, we implanted one array of 16 electrodes and two arrays of 32 electrodes in the LPFC, as well as two arrays of 16 electrodes in the FEF, for a total of 112 electrodes. The arrays consisted of platinum-iridium wires with either 200- or 400-µm separation, 1–5.5 mm long and with 0.5 MΩ of impedance, and they were arranged in 4×4 or 8×4 grids. Surgical procedures followed the following steps: 24 h prior to surgery, animals received a dose of dexamethasone to control inflammation during and after the surgery. They also received antibiotics (amoxicillin, 7–15 mg/kg, and Enrofloxacin, 5 mg/kg) for 8 d, starting 24 h before the surgery. During surgery, the scalp was incised and the muscles retracted to expose the skull. A craniotomy was performed (~2×2 cm). The dura mater was cut and removed from the craniotomy site. Arrays of electrodes were slowly lowered into the brain using a stereotaxic manipulator. Once all the arrays were secured in place, the arrays' connectors were secured on top of the skull using bone cement. A head-holder was also secured using bone cement. The piece of bone removed during the craniotomy was returned to its original location and secured in place using metal plates. The skin was sutured on top of the craniotomy site and stitched in place, avoiding any tension to ensure good healing of the wound. All surgeries were conducted using aseptic techniques under general anesthesia (isofluorane, 1–1.5% for maintenance). The depth of anesthesia was assessed by monitoring the heart rate and movement of the animal, and the level of anesthesia was adjusted as necessary. Analgesics were provided during postsurgical recovery, including a fentanyl patch (12.5 mg/2.5 kg, 24 h prior to surgery, removed 48 h after surgery) and meloxicam (0.2–0.3 mg/kg after removal of the fentanyl patch). Animals were not euthanized at the end of the study.

Recording techniques. Neural signals were initially acquired using a 128-channel and a 256-channel Plexon OmniPlex system (Plexon Inc., TX, USA) with a sampling rate of 40 kHz. The wideband signals were bandpass-filtered between 300 and 3,000 Hz. Following that, spikes were detected using an automated hidden Markov model-based algorithm for each channel[35]. Eye positions were obtained using an infrared-based eye-tracking device from SR Research Ltd. (Eyelink 1000 Plus). The behavioral task was designed on a standalone PC (stimulus PC) using the Psychophysics Toolbox[36] in Matlab (Mathworks, MA, USA). To align the neural and behavioral activity (trial epochs and eye data) for data analysis, we generated strobe words denoting trial epochs and performance (rewarded or failure) during the trial. These strobe words were generated on the stimulus PC and sent to the Plexon and Eyelink computers using the parallel port.

Microstimulation. For arrays positioned in the prearcuate region (FEF), we used standard electrical microstimulation to confirm that saccades could be elicited with low currents. These electrodes had a depth of 5.5 mm inside the sulcus and tapered to 1 mm away from sulcus. We conducted these microstimulation experiments after we finished our recording experiments. During the microstimulation experiment, each electrode implanted in the FEF was tested for its ability to evoke fixed-vector saccadic eye movements with stimulation at currents of 50 µA. Electrical microstimulation consisted of a 200-ms train of biphasic current pulses (1 ms, 300 Hz) with 0 interphase delays, delivered with a Plexon Stimulator (Plexon Inc., TX, USA). We mapped the saccade vector elicited via microstimulation at each electrode to verify that the electrodes were implanted in the FEF. Sites at which stimulation of 50 µA or less elicited eye movements at least 50% of the time, plus regions within 2–3 mm of these locations, were considered to be in the FEF[37].

Behavioral task. Each trial started with a mandatory period (500 ms) during which the animal fixated on a white circle at the center of the screen. While continuing to fixate, the animal was presented with a target (a red square) for 300 ms at any one of eight locations in a 3×3 grid. The center square of the 3×3 grid contained the fixation spot and was not used. The presentation of the target was followed by a delay of 1,000 ms, during which the animal was expected to maintain fixation on the white circle at the center. At the end of this delay, a distractor (a green square) was presented for 300 ms at any one of the seven locations (other than where the target was presented). This was again followed by a delay of 1,000 ms. At the end of the second delay the animal was given a cue (the disappearance of the fixation spot) to make a saccade towards the target location presented earlier in the trial. Saccades to the target location within 150 ms and continued fixation at the saccade location for 200 ms was considered a correct trial. An illustration of the task is shown in Figure 1a. One of the animals was presented with only seven of the eight target locations because of a behavioral bias in the animal.

Classical target selectivity. To investigate the responses of individual neurons to stimulus location, we computed the firing rate of neurons in non-overlapping 50-ms bins during all correct trials. The firing rates in each bin during the target period (50–350 ms after target onset), Delay 1 (400–1,200 ms after target onset), distractor period (1,350–1,650 ms after target onset), Delay 2 period (1,700–2,500 ms after target onset), and presaccadic (−200 to 200 ms with respect to saccade onset) were compared to the baseline (300 ms activity prior to the target period) using a two-sample t test ($P<0.05$) for each stimulus (target or distractor) location. A neuron was classified as responsive if the firing rate was significantly different from the baseline for 100 ms (two consecutive bins) for any of the stimulus locations. In addition, activities due to different stimulus (target or distractor) locations during each trial epoch (target, Delay 1, distractor, Delay 2, and response/saccade) were compared using one-way ANOVA ($P<0.05$). A selective neuron was one that exhibited responsiveness and a significant difference in activity between different stimulus locations in one task epoch for at least 100 ms (two consecutive bins). Depending on the task epoch in which the neuron exhibited selectivity and the stimulus (target or distractor) to which it was selective, the neuron was then classified into one of the following categories of selectivity: visual, delay, movement, or combinations of the above.

Percentage of explained variance. To compute the information in a neuronal population, we used the percentage of explained variance, ω^2, which computes the variance in firing rate for each neuron that can be explained by varying the spatial location of the stimulus. ω^2 was computed for every neuron in 100-ms bins with 50-ms steps. ω^2 is defined by

$$\omega^2 = \frac{SS_{\text{between-groups}} - (df \times MSE)}{SS_{\text{total}} + MSE} \times 100$$

where

$$SS_{\text{total}} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$SS_{\text{between-groups}} = \sum_{\text{group}}^{G} n_{\text{group}} (\bar{x}_{\text{group}} - \bar{x})^2$$

where df = degrees of freedom (6 or 7 depending on the animal), and MSE is the mean square error,

$$MSE = \sum_{i=1}^{n} (x_i - \bar{x}_{\text{group}})^2$$

We also balanced the number of trials in each group by stratifying the trials in each group to the lowest common number of trials across all groups. This was done to account for the bias in the calculation of PEV[38]. We repeated this step 1,000 times to form a distribution of PEV values at every timepoint for each neuron. To determine the time periods in which the PEV measure was significant, a randomization test was used. A null distribution of PEV values was created by shuffling the target labels associated with the trials (1,000 repetitions). This was repeated for each of the neurons, and only neurons that exhibited a z-score (mean and standard deviation for the z-score were computed from the null distribution) greater than 3 for at least one timepoint were used in computing the population averages for the LPFC and FEF (Fig. 1d). Figure 1d shows the mean and the 95th percentile confidence intervals of the mean PEV across all the neurons with $z>3$ in at least one time bin. To determine the time periods in which the PEV of the LPFC and FEF were significantly different, we compared the distribution of PEV values averaged across all significant neurons in the FEF and LPFC (Fig. 1d). For example, the PEV of LPFC neurons was considered significantly higher than FEF neurons if the 95th percentile distribution of PEV for LPFC neurons was higher than the 95th percentile distribution for FEF neurons.

Statistics. We considered two bootstrapped distributions to be significantly different if the 95th percentile ranges of the two distributions did not overlap. We also computed an estimated P value for this comparison using the following formula[39]:

$$\frac{1+X}{N+1}$$

where X represents the number of overlapping data points between the two distributions, and N represents the number of bootstraps. With this computation, and the $N=1{,}000$ bootstraps we used throughout the paper, two distributions with no overlap result in $P<0.001$, and two distributions with x% overlap result in $P \approx x/100$.

In addition to the estimated *P* value, we also computed the effect size of the comparison using a measure known as Hedges' *g*, computed using the following formula[40]

$$\left(1-\frac{3}{4(n_1+n_2)-9}\right)\times\left(\frac{\overline{x}_1-\overline{x}_2}{s'}\right)$$

where

$$s'=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

$\overline{x}$ refers to the mean of each distribution, *n* refers to the length of each distribution, and *s* refers to the standard deviation of each distribution.

No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications[11,21,22]. The majority of our analyses made use of nonparametric permutation tests and, as such, did not make assumptions regarding the distribution of the data. No randomization was used during the data collection, except in the selection of the target and distractor locations for each trial. Randomization was used extensively in the data analyzed to test for statistical significance. Data collection and analysis were not performed blind to the conditions of the experiments. No animals or data points were excluded from any of the analyzed. Please see additional information in the Life Sciences Reporting Summary.

Cross-temporal decoding. A decoder based on linear discriminant analysis (LDA) was built using the classify function in Matlab to predict the location of the target (Supplementary Software). To increase the number of neuronal responses used in this analysis, we pooled the responses across recording sessions to create a pseudopopulation of 256 neurons for the LPFC and 137 neurons for the FEF. We also analyzed recordings from single sessions (i.e., simultaneously recorded populations) and verified that they demonstrated the same trends (Supplementary Fig. 4). The instantaneous firing rate of each neuron (estimated with 100-ms windows with 50 ms of overlap) was converted to a *z*-score by normalizing to the mean and standard deviation of the instantaneous firing rates from 300 ms before target onset to target onset. The decoder was built to predict the seven target locations that were common to both the animals. Approximately 50% of the 3,143 correct trials from two animals (1,500 trials, with a uniform distribution of the seven different target locations) were grouped to build separate training datasets consisting of 1,500 pseudotrials for LPFC and FEF neurons. Subsets of the remaining correct trials, or 100% of the error trials, were used to build the testing dataset consisting of 100 pseudotrials. The training and testing data were denoised using principal components analysis (PCA) at every timepoint by reconstructing the data with the top *n* principal components that explained at least 90% of the variance. This step was done to avoid singular matrices when LDA was performed in Matlab and to reduce the noise in the data used to train and test the decoder. Initially, the target locations were predicted by training and testing the decoder on datasets from equivalent timepoints. The performance of the decoder, a measure of the information about target location in the population activity, was computed at each timepoint as a percentage of test trials in which the target location was predicted correctly. This process was repeated 1,000 times, with different subsets of correct trials used to constitute the training and test sets. This allowed us to obtain a distribution of performance for the decoders trained at each timepoint. At this point, we were ready to assess the stability of the population information across time. For each trained decoder, we also tested it with test sets from other timepoints throughout the trial. This gave us another 1,000 measures of performance at every other timepoint aside from the timepoint used to train the decoder. The average performance at every combination of training and testing timepoints is shown in heat maps (e.g., in Fig. 2a and other figures) displaying the results of cross-temporal decoding. To investigate the stability of the performance before and after the distractor, we first summarized the performance in the last 500 ms of Delay 1 (800–1,300 ms after target onset, shown in the dashed lines in the bottom left quadrant of Fig. 2a). This meant that we included the cross-temporal decoding performance of (i) a classifier trained on data from 800–900 ms after target onset and tested on subsequent windows up to the 1,200–1,300-ms window; (ii) a classifier trained on data from 1,200–1,300 ms and tested on preceding windows up to the 800–900-ms window; and (iii) all combinations of training and testing windows in between (Supplementary Fig. 2a,d). Since both the training data and the test data were from Delay 1, we called this performance LP_{11} or FP_{11} (for LPFC performance and FEF performance), where the first digit in the subscript indicates whether the training window was Delay 1 or Delay 2, and the second digit indicates whether the testing window was Delay 1 or Delay 2. To assess the performance of classifiers trained in Delay 1 and tested in Delay 2, we used the cross-temporal decoding performance of the same classifiers used to compute LP_{11} or FP_{11}, but this time we averaged the performance in a 500-ms window at the end of Delay 2 (i.e., 1,850–2,450 ms after target onset but before any presaccadic activity; shown in the dashed lines in the bottom right quadrant of Fig. 2a). This meant that we included the performance of (i) a classifier trained on the data from 800–900 ms after target onset and tested on windows starting from 1,850–1,950 ms and up to 2,350–2,450 ms; (ii) a classifier trained on the data from 1,200–1,300 ms and tested on the same windows as in (i); and (iii) all training windows in between, while the testing window was kept the same as in (i). Since the training data was from Delay 1 and the test data was from Delay 2, we called this performance LP_{12} or FP_{12}. The same procedure was used to compute LP_{22}, LP_{21}, FP_{22}, and FP_{21}. To check for statistical significance, we obtained the 95th percentile range for each of these values obtained through the 1,000 times we trained and tested different classifiers. If these ranges did not overlap, we considered the two performance values to be significantly different. This analysis was repeated every time we performed cross-temporal decoding on different subsets of neurons. In addition, we also subtracted the performance of LP_{22} from LP_{11}, LP_{12} from LP_{11}, and LP_{21} from LP_{22} (and similarly for the FEF) for each of the 1,000 repetitions, and we plotted the mean and standard error of the difference in the box plots in Supplementary Figure 2b,e. For significance, the 95th percentile range (two-sided) of this subtracted distribution was compared to 0.

State space analysis. To look at how the population response morphed from one delay period to another, we performed PCA on the responses in the same 1,500 pseudotrials that were used to train the cross-temporal decoding classifiers. The covariance matrix consisted of *z*-scores of each of the neurons in one area (LPFC or FEF) in one dimension, and in the other dimension, we included the responses from different time windows, different conditions, and different repetitions (i.e., trial-wise variance was also included). Our goal was to include the entirety of the space of Delay 1 (or Delay 2) responses and then perform PCA to identify the two axes that contained the largest variance. This allowed us to reduce the dimensionality of the LPFC population from 256 neurons to the two principal components with the largest variance and to reduce the FEF population from 137 neurons to two principal components. This made it possible for us to visualize the responses in a simple 2D plot. However, for the purposes of measuring shifts between Delay 1 and Delay 2 responses, we used the full PCA space instead of the reduced space to avoid underestimating the magnitude of the shifts. Although this PCA space did not correspond directly to the space used by the LDA classifier, it allowed us to gain some insight into the changes in the population responses between Delay 1 and Delay 2. The *z*-score of the instantaneous firing rate (described above) in 100-ms windows (with steps of 50 ms) for each delay period were used separately to define the space for dimensionality reduction. We used subsets from a separate 1,500 pseudotrials to perform the projections. For each target location, we selected 25 pseudotrials to compute an average *z*-score for each of the 256 LPFC neurons or the 137 FEF neurons. This allowed us to reduce some of the variability from individual trials. For visualization, the average *z*-scores for each 100-ms window from the full population were then projected onto the two principal components to obtain a point in the two-dimensional PCA space. To investigate what was happening in the last 500-ms window of Delay 1 and Delay 2, we averaged the points derived from the ten 100-ms windows to obtain the average position for those 500 ms. This procedure was repeated to obtain 50 points for each target location. To obtain the intracluster distance, we computed all the pairwise Euclidean distances in the full PCA space (normalized by dividing by the number of dimensions, i.e., the number of neurons) for all the points in each target location and then averaged the distances across target locations. To obtain the intercluster distance, we averaged the 50 points in each target location to obtain a point in the full PCA space for each target location. We then computed all the pairwise distances between these points and averaged them to obtain the intercluster distance. By repeating the analysis 1,000 times with different sets of 25 pseudotrials, we were able to compute the mean and the range of distances spanned by 95% of the repetitions. We computed the intra- and intercluster distances for Delay 1 and Delay 2 clusters in Delay 1 or Delay 2 space correspondingly. The inter- and intracluster distances between Delay 1 and Delay 2 were considered significantly different if the 95th percentile distributions of each distance (Delay 1 or Delay 2) did not overlap. To quantify the shift in the cluster locations, we first computed the cluster center by taking the average of the points in each cluster. We then computed the Euclidean distances between each of the cluster centers in Delay 1 (i.e., the distances between the blue clusters and red clusters for each target location in Fig. 3a,e), as well as the distances between corresponding cluster centers in Delay 2 (Fig. 3b,f). The average distance was then computed from these distances. We repeated this 1,000 times to obtain a distribution of average distances. The 95th percentile range is depicted with error bars in Figure 3d,h. To test for statistical significance, we created 1,000 pairs of subsampled Delay 1 responses in Delay 1 space for each of the locations and computed the average intracluster shift across locations. We repeated this for Delay 2 responses in Delay 2 space. The 97.5th percentile range of these chance distributions are shown in the dashed lines in Figure 3d,h. This analysis was done whenever we computed the shift in cluster centers.

Distractor specificity in target code. Aside from looking at cross-temporal decoding, we also looked to see if there was any evidence that the target information maintained in Delay 2 was specific to the distractor location. We did this by training classifiers on Delay 2 data from trials in which the target was located at any of the seven possible locations but the distractor was located

at just one of the locations. These classifiers were then tested to see if they could generalize to decode target location in trials in which the distractors were presented at other locations. In this analysis, we used all the neurons within each area, including those with target–distractor NMS. The results for the LPFC and FEF are shown in Supplementary Figure 6a,d. Each heat map shows the target decoding performance (averaged over 1,000 iterations) on test trials for a classifier trained on trials with distractors at only one location. For example, the heat map shown in the top left indicates the test results for a classifier trained on trials in which the distractor was presented at the top left location. The test results for different target locations in trials in which the distractor was presented at the same top left location are shown in the first column, with the top left location shown in the bottom row. This corresponds to the within-condition performance. As the target and distractor never appeared at the same location in our experiment, the classifier was never trained on any trials in which the target appeared at the top left location. As a result, we did not test the classifier on this location, and this is why the entire bottom row is blank. The remaining columns illustrate the test results for the classifier in trials where the distractor was in different locations, corresponding to the cross-condition performance. The blank squares along the diagonal are again due to the fact that the location of targets and distractors never overlapped in our experiment. If the morphed population information was specific to the distractor location, we would expect significant differences across the columns in a row. We did this by comparing the bootstrapped performance distributions for all possible pairs within each row and checking whether the 2.5th and 97.5th percentile range overlapped.

We also averaged the classifier performance across target locations and plotted the average performance and the 2.5th and 97.5th percentiles in the boxplots in Supplementary Figure 6b,e. The performance in within-condition trials are labeled *w*, while those for cross-condition trials are labeled *c*. If the 95th percentile range overlapped for all possible pairs, it would indicate that training the classifier on the responses at one distractor location did not significantly affect its classification performance when tested with responses at other distractor locations. We also directly compared performances within conditions and across conditions in Supplementary Figure 6c,f for each of the seven types of classifiers. We subsampled the cross-condition performance to make sure we had the same number of measurements as those in the within-condition performance and checked for overlap in the 95th percentile range. However, one caveat for the above analysis is the large error bars in the plots indicating the 95th percentile range, which were the result of the small number of trials for some target–distractor combinations. As this number of trials has to be reduced further into training and test trials, this greatly reduces the number of trials available for each combination.

As an alternative to splitting the data into training and testing sets, we looked for evidence of distractor specific code morphing, using the state space analysis shown in Figure 3. Using the same state space as the analysis shown in Figure 3, we used a cluster-separation measure (LDA S-ratio[41], which computes the variance between classes divided by variance within class) to quantify clustering in the Delay 2 responses when the points were grouped by distractor location. To maximize the number of trials we had for each target–distractor combination, we selected the four target locations for which we had the largest numbers of trials. In addition, we used the single-session recordings instead of the pseudopopulation data, as the latter required us to remove trials to match the lowest number of trials across sessions. The results are shown in Supplementary Figure 7. The cluster separations (measured in the full state space) for each group of points representing different distractor locations were computed, and the average and standard errors are shown in Supplementary Figure 7b,c,e,f. As a control, we randomly grouped the points into seven clusters and computed the average cluster separation. We repeated this 1,000 times and obtained the 97.5th percentile in cluster separation (dashed lines). As another control, we performed the same analysis on the responses in Delay 1, which are also plotted in the same plots. Since the distractors were presented after Delay 1, the responses in Delay 1 should not exhibit any clustering by distractor location.

Error trials. Due to an unequal distribution of error trials across different locations, we only analyzed the error trials from four target locations. We only included those locations in the analysis where there were at least six error trials in every session. For calculating the shift caused by chance in Figure 4e,j, we created 1,000 pairs of subsampled Delay 1 responses in Delay 1 space for each of the four locations and computed the average intracluster shift for error trials across locations. We repeated this for Delay 2 error trials in Delay 2 space. The 97.5th percentile range of these chance distributions are shown in the dashed lines in Figure 4e,j.

Comparing effect sizes in neurons with NMS. After performing two-way ANOVA to identify neurons with NMS, we used the *F*-statistic returned by the two-way ANOVA to compare the effect sizes of neurons with NMS in the LPFC and FEF. As we found 13 neurons with NMS in the FEF and 54 in the LPFC, we subsampled the LPFC population to produce 13 neurons at a time to compute the mean and compare against the mean of the FEF population. We performed the subsampling 1,000 times and obtained the 95th percentile range for the mean computed from 13 LPFC cells.

Role of different neurons in code morphing. To investigate the role of different types of neurons in mediating code morphing in the LPFC, we first selected 27 neurons with NMS, 27 neurons with LMS, and 27 neurons with CS. The cross-temporal decoding was then performed on the following subpopulations: (i) without neurons with NMS, (ii) without neurons with LMS, and (iii) without neurons with CS. We also performed a separate analysis on homogeneous populations of 27 neurons with NMS, 27 neurons with LMS, and 27 neurons with CS.

Cluster distances for LPFC neurons with different selectivity. To look at how different LPFC neurons contribute to the changes in the intercluster and intracluster distances, we created a population consisting of equal numbers of neurons with NMS, neurons with LMS, and neurons with CS. As the smallest number of neurons was 27 for the neurons with LMS, we randomly selected 27 of the 64 neurons with NMS and 27 of the 30 neurons with CS. We then looked at code morphing when one of the neuron types was excluded (i.e., we were left with 54 neurons), as well as when only one of the neuron types was included (i.e., only 27 neurons used). We used the same 1,500 pseudotrials to create the PCA space, but this time with just 54 or 27 neurons. For illustration, we again selected the two principal components with the largest variances and projected responses from a separate set of 1,500 trials onto the two principal components to obtain points in the two-dimensional PCA space (Supplementary Fig. 10). However, the shifts shown in Figure 6d,h were distances measured for the full 54 and 27 dimensions, respectively.

Injection of neurons with NMS into the FEF. To demonstrate the ability of neurons with NMS to morph the population code, we added 20 of the most nonlinear neurons with NMS from the LPFC to the population of 137 FEF neurons. This allowed us to increase the percentage of neurons with NMS to match the percentage found in the LPFC (25%). It also allowed us to alter the average *F*-statistic for the modified population of neurons with NMS to fall within the 2.5th and 97.5th percentile range found in the LPFC. The results of the cross-temporal decoding and the state space analysis from this modified FEF population are shown in Supplementary Figure 11a–d. With the addition of the 20 neurons with NMS to the FEF, the code morphing increased to significantly above chance in both the Delay 1 and Delay 2 spaces. This meant that by adding the neurons with NMS, the new FEF population gained the ability to morph the code, much like the LPFC population.

Decoding with FEF-like NMS neurons in the LPFC. To demonstrate the role of neurons with NMS in morphing the population code, we attempted to match the effect sizes of the neurons with NMS in the LPFC with those in the FEF (Fig. 5f). We did this by removing 23 LPFC neurons with $F > 6$, which was the largest value found in the FEF population. This also reduced the proportion of neurons with NMS to 17.6%, close to the 19% we observed in FEF. The results of the state space analysis from this modified LPFC population are shown in Supplementary Figure 11e–h. The code morphing in both Delay 1 and Delay 2 spaces dropped below significance, providing further evidence that the neurons with NMS with large interaction effects (which were unique to the LPFC) were primarily responsible for code morphing.

Neurons with target–distractor nonlinear mixed selectivity. To identify a nonlinear target–distractor mixed selective neuron, we computed the firing rates of neurons in 50-ms bins for different target–distractor pairs during all correct trials. This accounted for 56 or 42 different pairs of target–distractor locations (8 or 7 target locations and 7 or 6 distractor locations associated with each target location) for Monkey A or B, respectively. Each category of target–distractor pair consisted of about 6–10 correct trials for each neuron. Due to the limited number of trials for each category, we compared the trial-averaged activity for each category to a null distribution computed from the trials with the same target location but with distractors presented at any of the locations except the one associated with the category in comparison. For example, the trial-averaged firing rate of a neuron for trials presented with the target at Location 1 (left corner) and the distractor at Location 2 (top center) was compared to a null distribution generated from trials with the target at Location 1 and the distractor at every location other than Location 2. A null distribution with 100 bootstraps was computed, with each bootstrap representing the average activity from the same number of trials as the target–distractor pair in question. A neuron was classified as nonlinear mixed selective using a surrogate test if the trial-averaged activity of one or more target–distractor pair was different from the 2.5th and 97.5th percentile of the null distribution for 100 ms (two consecutive bins) during the distractor and Delay 2 periods.

Decoding without neurons with target–distractor nonlinear mixed selectivity. We investigated the role of the neurons with target–distractor nonlinear mixed selectivity in mediating the morphing in population code by excluding these neurons ($n = 31$) from our pseudopopulation in the LPFC. The results of the state space analysis from this modified LPFC population are shown in Supplementary Figure 14a–d. Excluding the neurons with tdNMS did not appear

to affect the code morphing in either the Delay 1 or Delay 2 spaces, suggesting that in our task, the neurons with tdNMS did not play a big role in the code morphing in the LPFC.

Receptive field comparison between LPFC and FEF. To investigate whether differences in receptive field (RF) sizes in the LPFC and the FEF can account for the differences in population code morphing we observed, we quantified receptive field sizes by counting the number of target locations that exhibited responses that were significantly different from the baseline. This was done using the same measure used to quantify classical target selectivity. To perform a statistical comparison, we subsampled the population of LPFC and FEF neurons to create 1,000 bootstrapped populations of 100 neurons each. We computed the histogram of the receptive field size for each bootstrapped population and then computed a weighted average ($N_{neurons} \times$ size of RF) for each bootstrapped population. The 95th percentile range of the weighted average was compared for the two bootstrapped distributions to see if they overlapped.

Data availability. The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Code availability. A code package for performing the cross-temporal decoding is available at https://github.com/aishu1803/Code-for-Parthasarathy-et-al.

References

35. Herbst, J. A., Gammeter, S., Ferrero, D. & Hahnloser, R. H. Spike sorting with hidden Markov models. *J. Neurosci. Methods* **174**, 126–134 (2008).
36. Brainard, D. H. The psychophysics toolbox. *Spat. Vis* **10**, 433–436 (1997).
37. Bruce, C. J., Goldberg, M. E., Bushnell, M. C. & Stanton, G. B. Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. *J. Neurophysiol* **54**, 714–734 (1985).
38. Buschman, T. J., Siegel, M., Roy, J. E. & Miller, E. K. Neural substrates of cognitive capacity limitations. *Proc. Natl. Acad. Sci. USA* **108**, 11252–11255 (2011).
39. Ojala, M. & Garriga, G. C. Permutation tests for studying classifier performance. *J. Mach. Learn. Res.* **11**, 1833–1863 (2010).
40. Hedges, L. V. Distribution theory for Glass's estimator of effect size and related estimators. *J. Educ. Behav. Stat.* **6**, 107–128 (1981).
41. Fisher, R. A. The use of multiple measurements in taxonomic problems. *Ann. Hum. Genet* **7**, 179–188 (1936).

nature research

Corresponding author(s): Camilo Libedinsky & Shih-Cheng Yen

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

We performed our experiments on two monkeys. We recorded 256 neurons from lateral Prefrontal cortex and 137 neurons from Frontal Eye Fields across 8 recording sessions. The sessions were chosen based on the number of trials performed by the monkey in the session.

2. Data exclusions

Describe any data exclusions.

Sessions that had less than 200 trials were excluded from this work.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Experiments were carried out in 2 animals. All results replicated successfully in both animals. No differences were observed in any analysis between animals.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Not applicable, since we do not have different experimental groups.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

N/A

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	All analyses were performed in MATLAB.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	N/A

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	N/A

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	N/A
b. Describe the method of cell line authentication used.	N/A
c. Report whether the cell lines were tested for mycoplasma contamination.	N/A
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	N/A

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	We used non-human primates (Macaca fusicularis) for this work.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	N/A