

# Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles

Matthew L. Leavitt<sup>a,b,1</sup>, Florian Pieper<sup>c</sup>, Adam J. Sachs<sup>d</sup>, and Julio C. Martinez-Trujillo<sup>a,b,e,f,g,1</sup>

<sup>a</sup>Department of Physiology, McGill University, Montreal, QC, H3G 1Y6, Canada; <sup>b</sup>Department of Physiology and Pharmacology, University of Western Ontario, London, ON, N6A 5B7, Canada; <sup>c</sup>Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany; <sup>d</sup>Division of Neurosurgery, Ottawa Hospital Research Institute, University of Ottawa, Ottawa, ON, K1N 4E9, Canada; <sup>e</sup>Robarts Research Institute, University of Western Ontario, London, ON, N6A 5B7, Canada; <sup>f</sup>Department of Psychiatry, University of Western Ontario, London, ON, N6A 5B7, Canada; and <sup>g</sup>Brain and Mind Institute, University of Western Ontario, London, ON, N6A 5B7, Canada

Edited by Ranulfo Romo, Universidad Nacional Autónoma de México, Mexico City, D.F., Mexico, and approved February 13, 2017 (received for review December 5, 2016)

Neurons in the primate lateral prefrontal cortex (LPFC) encode working memory (WM) representations via sustained firing, a phenomenon hypothesized to arise from recurrent dynamics within ensembles of interconnected neurons. Here, we tested this hypothesis by using microelectrode arrays to examine spike count correlations ( $r_{sc}$ ) in LPFC neuronal ensembles during a spatial WM task. We found a pattern of pairwise  $r_{sc}$  during WM maintenance indicative of stronger coupling between similarly tuned neurons and increased inhibition between dissimilarly tuned neurons. We then used a linear decoder to quantify the effects of the high-dimensional  $r_{sc}$  structure on information coding in the neuronal ensembles. We found that the  $r_{sc}$  structure could facilitate or impair coding, depending on the size of the ensemble and tuning properties of its constituent neurons. A simple optimization procedure demonstrated that near-maximum decoding performance could be achieved using a relatively small number of neurons. These WM-optimized subensembles were more signal correlation ( $r_{signal}$ )-diverse and anatomically dispersed than predicted by the statistics of the full recorded population of neurons, and they often contained neurons that were poorly WM-selective, yet enhanced coding fidelity by shaping the ensemble's  $r_{sc}$  structure. We observed a pattern of  $r_{sc}$  between LPFC neurons indicative of recurrent dynamics as a mechanism for WM-related activity and that the  $r_{sc}$  structure can increase the fidelity of WM representations. Thus, WM coding in LPFC neuronal ensembles arises from a complex synergy between single neuron coding properties and multidimensional, ensemble-level phenomena.

working memory | prefrontal cortex | noise correlations | macaque | decoding

To interact with a complex, dynamic environment, organisms must be capable of maintaining and manipulating information that is no longer available to their sensory systems. This capability, when applied transiently (i.e., for milliseconds to seconds), is referred to as working memory (WM) (1)—a hallmark of intelligence and a crucial component of goal-directed behavior (2). In 1949, Hebb postulated that sustained neuronal activity in the absence of stimulus input could serve as the neural substrate for WM (3). Fuster and Alexander later discovered neurons in the lateral prefrontal cortex (LPFC) of monkeys that exhibited sustained firing during WM tasks (4). Subsequent neurophysiological studies have corroborated that neuronal activity in the LPFC and other regions can represent WM for visual-mnemonic space (5–7), as well as nonspatial visual features (8–10).

Electrophysiological studies of spatial WM have traditionally relied on recording from one neuron or a few neurons simultaneously (10). However, the neuronal computations that underlie sophisticated behaviors such as WM require the coordinated activity of many neurons within and across brain networks (11). We currently lack a clear understanding of how single neuron coding properties scale to neuronal ensembles. Can the properties of an ensemble be predicted by aggregating the individually and in-

dependently measured properties of its constituent neurons? The answers to this question and related questions hinge on how ensembles are affected by phenomena that emerge from interactions between neurons.

The sustained activity presumed to underlie WM maintenance is thought to be achieved by increasing the strength of recurrent excitation and lateral inhibition between neurons within an ensemble (12–18). These dynamics should modify patterns of correlated firing between neurons in a manner dependent on differences in their tuning properties. Such a pattern can be quantitatively characterized by two measurements: The first is signal correlation ( $r_{signal}$ ), the similarity of two neurons' responses to a set of different stimuli or experimental conditions. The second is spike count (or noise) correlations ( $r_{sc}$ ), the similarity in the variability of two neurons' responses to the same stimulus or experimental condition (19).

Given a fixed ensemble of neurons (and thus a constant  $r_{signal}$  structure), changes in  $r_{sc}$  can have profound effects on information coding (19–24). For example, spatial attention improves neural coding in the visual cortex primarily by reducing  $r_{sc}$  (25–27). Another study reported that increased  $r_{sc}$  improved perceptual discrimination in macaque area S2 (28). These results are difficult to extend to WM coding in the LPFC. Furthermore, there are relatively few studies investigating  $r_{sc}$  in the LPFC (21, 27, 29–33); and only one of these studies directly examined the effects of  $r_{sc}$  on information coding (27). Prior results examining pairwise correlations are also difficult to extrapolate to larger neuronal ensembles, which have a complex, multidimensional  $r_{sc}$  structure that cannot be characterized by pairwise measurements alone (20).

## Significance

The working memory (WM)-related activity in the primate prefrontal cortex (PFC) is hypothesized to arise from the structure of the network in which the neurons are embedded. Recent studies have also shown that it is difficult to predict the properties of neuronal ensembles from the properties of individually examined neurons. By recording the activity of neuronal ensembles in the macaque PFC, we found evidence supporting the network origins of WM activity and discovered features of WM coding in neuronal ensembles that were inaccessible in prior single neuron studies. Most notably, we found that correlated firing rate variability between neurons (i.e., noise correlations) can improve WM coding and that neurons not selective for WM can improve WM coding when part of an ensemble.

Author contributions: M.L.L., F.P., A.J.S., and J.C.M.-T. designed research; M.L.L. performed research; M.L.L. analyzed data; and M.L.L. and J.C.M.-T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence may be addressed. Email: julio.martinez@robarts.ca or matthew.leavitt@mail.mcgill.ca.

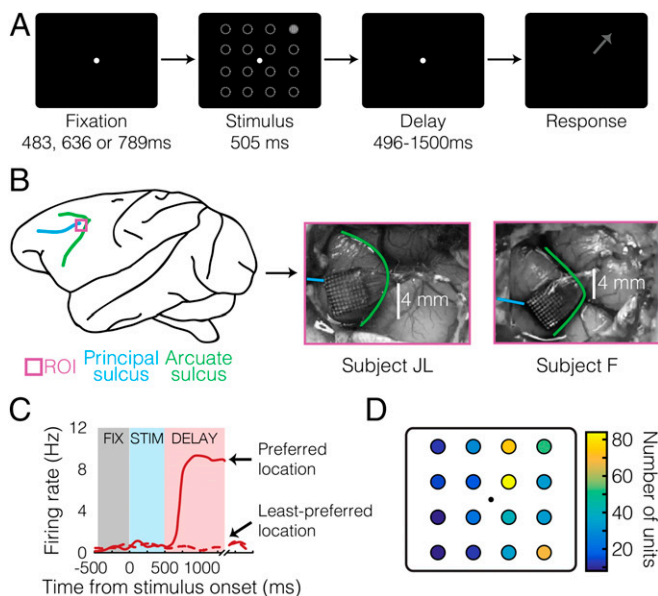
This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1619949114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1619949114/-DCSupplemental).

Currently it remains unknown whether and how  $r_{sc}$  structure modulates the fidelity of WM coding in LPFC neuronal ensembles.

We used microelectrode arrays (MEAs) to record from neuronal ensembles in the LPFC of two monkeys while they performed an oculomotor delayed-response task and assessed ensemble information content using a linear decoder. We found that  $r_{sc}$  varied as a function of  $r_{signal}$  during WM maintenance in a manner predicted by a recurrent excitation and lateral inhibition scheme. Using all simultaneously recorded neurons, the decoder could reliably predict which of 16 locations was being remembered. We also devised procedures to systematically investigate how WM coding varies across the “configuration space” of potential neuronal ensembles. Removing the  $r_{sc}$  structure could increase or decrease the information content of neuronal ensembles across the configuration space. However, the intrinsic  $r_{sc}$  structure improved WM coding in smaller neuronal subensembles of neurons optimized for WM representation. These optimized ensembles had a stereotyped  $r_{signal}$  distribution, with peaks at zero and extreme negative values, and spanned farther across the cortical surface than predicted by the statistics of the full population of recorded LPFC units. Finally, we observed individual units that did not encode WM in isolation (“nonselective” neurons) but that still contributed to WM coding when part of an ensemble by altering the  $r_{sc}$  structure.

## Results

Two adult male monkeys (*Macaca fascicularis*) (subjects “JL” and “F”) performed an oculomotor delayed-response task (Fig. 1A) while we recorded from neuronal ensembles in the left LPFC area 8A, anterior to the arcuate sulcus and posterior to the principal, using chronically implanted 96-channel microelectrode arrays (Fig. 1B). The neural correlates of WM for spatial locations have been extensively documented in this brain region (10). The target stimulus could appear at any one of 16 possible locations, arranged in a uniformly spaced  $4 \times 4$  grid around a central fixation point. We collected spike data from a total of 545 single units and multiunits across 12 recording sessions, out of which 417 (76%) exhibited



**Fig. 1.** Task, method, and single-cell data. (A) Overview of oculomotor delayed-response task. The arrow represents the correct saccade direction. The dashed circles indicate potential cue locations and are shown for illustrative purposes only and are not present in the task. (B) Array implantation sites and anatomical landmarks in both subjects. (C) Example delay-selective neuron. (D) Distribution of delay-selective units' preferred locations. FIX, fixation; ROI, region of interest; STIM, stimulus.

sustained activity and selectivity during the delay epoch ( $P < 0.05$ , Kruskal–Wallis; firing rate  $\times$  location) (*Materials and Methods*). We included both multiunits and single units in our analyses, as in similar previous studies (21, 25, 27, 34). A unit's preferred location during a given epoch was defined as the location that elicited the largest response averaged over that epoch (Fig. 1C and D). Subjects made incorrect choices about the stimulus location in  $<1\%$  of completed trials. Only correct, completed trials were included for analysis.

**Task-Related Modulation of Spike Count Correlations.** We computed spike count correlations ( $r_{sc}$ ) between pairs of neurons (pairwise  $r_{sc}$ ) (*Materials and Methods*) during the fixation, stimulus, and delay epochs.  $r_{sc}$  can covary with firing rate (21, 35) so, to ensure that differences in  $r_{sc}$  across epochs were not confounded by differences in firing rates, we implemented a distribution-matching procedure (*Materials and Methods*). We replicated two findings from previous studies: Mean pairwise  $r_{sc}$  was significantly above zero in each task epoch (Fig. 2A) ( $P < 0.005$  for all epochs, bootstrap test) (*Materials and Methods*); and  $r_{sc}$  varied as a function of tuning similarity, which we quantified as signal correlation between pairs of neurons during the delay epoch ( $r_{signal}$ ) (*Materials and Methods*) (Fig. 2B) (18, 29–33). Specifically, we found that the median pairwise  $r_{sc}$  was consistently larger for similarly tuned neuron pairs (defined as  $r_{signal} > 0.25$ ) compared with dissimilarly tuned pairs (defined as  $r_{signal} < -0.25$ ) ( $P < 0.001$  for all epochs, bootstrap test) (*Materials and Methods*). We also found that mean pairwise  $r_{sc}$  was greater during the fixation and delay epochs compared with the stimulus epoch (Fig. 2A) ( $P < 0.001$  for both fixation vs. stimulus and delay vs. stimulus, bootstrap test). Most importantly, we found that the relationship between  $r_{sc}$  and  $r_{signal}$  changed across task epochs (Fig. 2B); specifically, median pairwise  $r_{sc}$  for similarly tuned neurons was larger during the delay epoch than during the fixation and stimulus epochs ( $P < 0.001$  for both comparisons, bootstrap test) (Fig. 2C), and median pairwise  $r_{sc}$  between dissimilarly tuned neurons was lower during the stimulus and delay epochs than during the fixation epoch ( $P < 0.001$  for both comparisons, bootstrap test) (Fig. 2C). These results indicate that WM maintenance modifies pairwise  $r_{sc}$  in the LPFC in a manner consistent with a recurrent excitation, lateral inhibition scheme (12–17, 36).

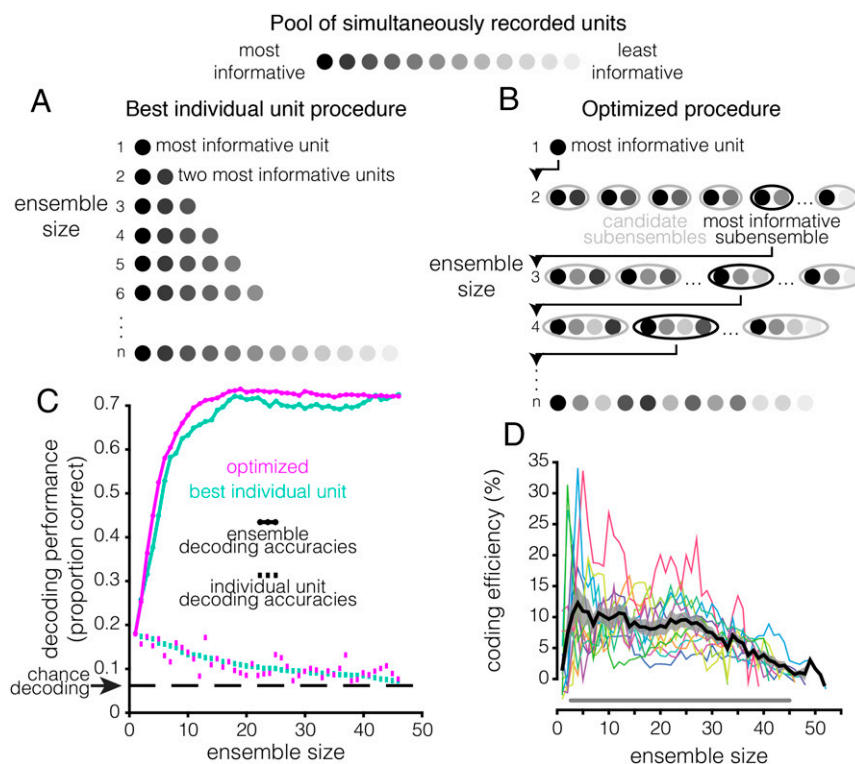
**Quantifying Information Content in Neuronal Ensembles Using Linear Decoders.** Pairwise measurements of  $r_{sc}$  are insufficient for predicting the effects of  $r_{sc}$  structure on ensemble information in large, multidimensional ensembles with heterogeneous tuning (20). Furthermore, analytical methods for determining the effects of  $r_{sc}$  structure on information content can be complicated to calculate for large stimulus sets and can also be inaccurate unless applied to data consisting of hundreds of trials per stimulus (20, 37). Linear decoders are demonstrably well-suited for extracting low-dimensional representations from high-dimensional neuronal ensemble data and for directly assessing the impact of  $r_{sc}$  structure on ensemble information content and thus offer a pragmatic solution to the issues of dimensionality and correlated variability (20, 38).

Previous studies have decoded the identity of stimuli maintained in spatial (39, 40) and nonspatial WM (8, 39) in pseudo-populations of LPFC neurons, typically using sets of 2 to 8 unique stimuli. We were able to reliably decode which of 16 target locations was being held in WM during the delay epoch by applying a linear support vector machine (SVM) (*Materials and Methods*) to simultaneously recorded ensemble data (Fig. S14) (max = 77%; mean across sessions = 52%).

Examining ensembles consisting of every simultaneously recorded neuron and/or only tuned neurons is a standard practice in neurophysiology. However, this practice assumes that all of the examined neurons contribute to coding, an assumption difficult to verify. It is possible that a subset of the recorded neurons can represent nearly as much information as the entire ensemble and that such a subset could form a “unit” of information coding that is







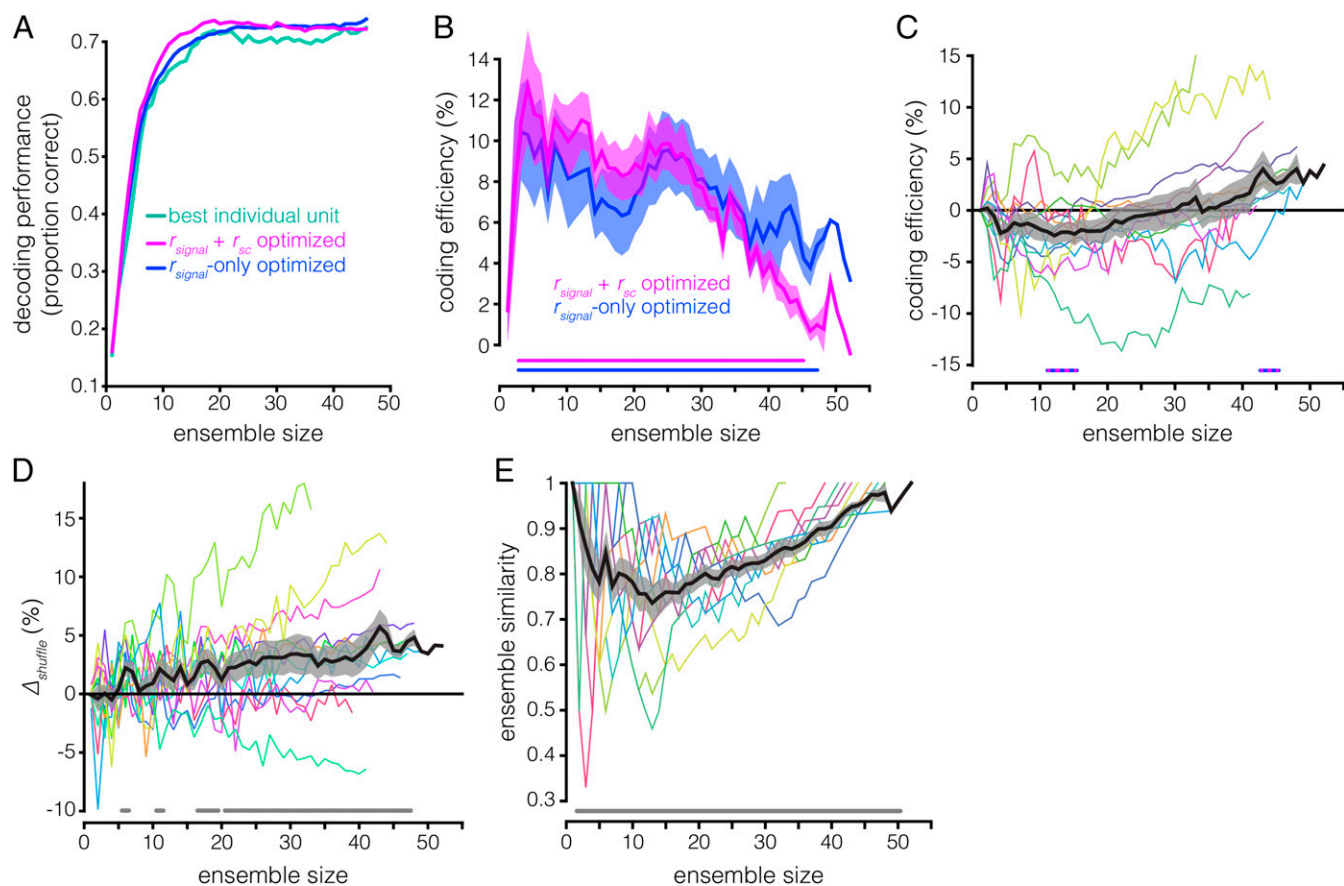
**Fig. 3.** Accounting for between-neuron phenomena increases ensemble efficiency. Visualization of the (A) best individual unit ensemble construction procedure and (B) optimized ensemble construction procedure. Each circle represents a unit, and the shading represents that unit's information content, as assessed using the decoder. (C) Decoding results for the best individual unit (teal) and optimized procedures (violet), applied to a single example session. The continuous line plot with circular markers shows the ensemble decoding accuracy (y axis) as a function of size (x axis). The square markers at the bottom of the plot denote the decoding accuracy (y axis) of the individual unit added to the ensemble at a given size (x axis). Both methods yield identical results for ensembles of the maximum size because these ensembles are identical; they consist of every simultaneously recorded unit in the session (i.e., the full ensemble). (D) Coding efficiency of the optimized method relative to the best individual unit method (y axis) as a function of ensemble size (x axis). Coding efficiency is quantified as  $[(accuracy_{optimized}/accuracy_{best\ individual\ unit}) - 1] \times 100$ . Colored lines are values for individual sessions. The thick black line is the across-session mean, and the gray shaded area is the SEM. The gray line running along the bottom indicates ensemble sizes for which the optimized method is significantly more efficient than the best individual unit method ( $P < 0.05$ , paired  $t$  test, Hochberg-corrected).

individual unit method requires  $\sim 33\%$  of the units ( $\sim 14$  units). In “random” ensembles—ensembles generated by randomly subsampling  $n$  units from a given recording session— $\sim 85\%$  of the units are necessary to reach 95% of maximum decoding accuracy. These results demonstrate that neuronal ensembles in the LPFC encode more information than single neurons, that the most informative ensembles are not necessarily composed of the most informative individual units, and that a relatively small subset of neurons can represent nearly as much WM information as the full recorded population.

**Effects of  $r_{sc}$  and  $r_{signal}$  Structures on WM Coding Efficiency.** To dissociate the effects of the  $r_{sc}$  and  $r_{signal}$  structures on WM coding efficiency in the optimized ensembles, we constructed new ensembles using the optimized procedure on firing rate data from which the  $r_{sc}$  structure had been removed via shuffling; the classifier was trained and tested on shuffled data for each ensemble size. We then compared the information content of these “ $r_{signal}$ -only” ensembles with the information content of ensembles generated using the original,  $r_{sc}$  structure-intact data, which we now refer to as the “ $r_{signal} + r_{sc}$ ” ensembles. The results for all three methods (best individual unit,  $r_{signal} + r_{sc}$ -optimized, and  $r_{signal}$ -only-optimized) applied to an example session are compared in Fig. 4A. The  $r_{signal}$ -only ensembles contain significantly more WM information than the best individual unit ensembles across sizes ranging from 2 to 47 neurons ( $P < 0.05$ , paired  $t$  test, Hochberg-corrected) (Fig. 4B). However, the effect of the  $r_{sc}$  structure is variable: The  $r_{signal} + r_{sc}$  ensembles are more efficient than the

$r_{signal}$ -only ensembles at smaller ensemble sizes whereas this effect inverts at larger ensemble sizes ( $P < 0.05$  for ensemble sizes of 11 to 15 and 43 to 45 neurons, paired  $t$  test, Hochberg-corrected) (Fig. 4B and C). These changes in WM coding efficiency effected by the  $r_{sc}$  structure can reach  $\pm 15\%$  across different recording sessions and ensemble sizes (Fig. 4C), enough to double (or nullify) efficiency increases afforded by the  $r_{signal}$  structure alone. These results indicate that the  $r_{sc}$  structure significantly impacts WM coding and can do so in a manner that varies non-monotonically with ensemble size. These results cannot be ascribed to idiosyncrasies of the SVM decoder because repeating the same analyses using logistic regression yields similar results (Fig. S3). We also found a similar—though less consistent—effect during stimulus presentation, with considerably greater session-to-session variability (Fig. S4).

It is possible that the observed effects of the  $r_{sc}$  structure on WM coding are simply a property of an ensemble's size, regardless of whether the ensemble is optimized for WM representation. To resolve this ambiguity, we compared the decoding performance of the random ensembles in which  $r_{sc}$  structure was intact vs. shuffled (Fig. 4D). We found that shuffling out the  $r_{sc}$  structure significantly improved decoding in most ensembles of six or more units ( $P < 0.05$ , paired  $t$  test, Hochberg-corrected) and that the magnitude of the decoding improvement was robustly and significantly correlated with the size of the ensemble in 8 out of 12 recording sessions (Spearman's  $\rho \geq 0.53$ ;  $P < 0.001$ ) (Fig. S5A). Although the  $r_{sc}$  structure seems to consistently impair decoding at the largest ensemble sizes (Fig. S5B), these results demonstrate that WM



**Fig. 4.** Effects of  $r_{sc}$  structure on ensemble coding efficiency and composition. (A) Decoding accuracy (y axis) as a function of ensemble size (x axis) for the best individual unit (teal),  $r_{signal} + r_{sc}$  (violet), and  $r_{signal}$ -only (blue) methods for the same example session as in Fig. 3C. Note that, for the  $r_{signal}$ -only ensembles, the classifier was trained and tested on  $r_{sc}$ -shuffled data whereas, for the  $r_{signal} + r_{sc}$  and best individual unit ensembles, the classifier was trained and tested on  $r_{sc}$ -intact data. (B) Coding efficiency of  $r_{signal} + r_{sc}$  ensembles and  $r_{signal}$ -only ensembles, relative to the best individual unit ensembles (y axis), as a function of ensemble size (x axis). The violet line running along the bottom indicates ensemble sizes for which the  $r_{signal} + r_{sc}$  ensembles are significantly more efficient than the best individual unit ensembles ( $P < 0.05$ , paired t test, Hochberg-corrected); the blue line is similar, but for  $r_{signal}$ -only ensembles vs. best individual unit ensembles. Note that the coding efficiency of  $r_{signal} + r_{sc}$  ensembles relative to best individual unit ensembles was previously shown in Fig. 3D. (C) Coding efficiency of  $r_{signal}$ -only ensembles relative to  $r_{signal} + r_{sc}$  ensembles; similar to Fig. 3D. A positive value indicates that shuffling out the  $r_{sc}$  structure improves decoding. The striped blue and violet lines running along the bottom indicate ensemble sizes for which the efficiency of  $r_{signal} + r_{sc}$  ensembles and  $r_{signal}$ -only ensembles are significantly different ( $P < 0.05$ , paired t test, Hochberg-corrected). (D) Decoding performance of  $r_{sc}$ -shuffled vs.  $r_{sc}$ -intact ensembles ( $\Delta_{shuffle}$ , y axis) as a function of ensembles size (x axis) for random ensembles. Ensembles were generated by randomly subsampling  $n$  units from the full recorded population in a given session. The gray lines running along the bottom indicate ensemble sizes for which the  $r_{sc}$ -shuffled vs.  $r_{sc}$ -intact ensembles are significantly different ( $P < 0.05$ , paired t test, Hochberg-corrected). (E) Similarity between  $r_{signal} + r_{sc}$  ensembles and  $r_{signal}$ -only ensembles (y axis) as a function of ensemble size (x axis). Ensemble similarity is quantified as the proportion of units common to the two ensembles for a given size. Note that ensemble similarity is 1 for ensembles of size  $n = 1$ , and for the largest ensemble size in a given session, because both ensemble-building procedures begin with the same unit, and the largest ensemble in each session consists of every simultaneously recorded unit in that session. The gray line running along the bottom indicates ensemble sizes for which the similarity of the  $r_{signal} + r_{sc}$  ensembles and  $r_{signal}$ -only ensembles is significantly less than 1 ( $P < 0.05$ , z-test of proportion, Hochberg-corrected).

coding in a neuronal ensemble consisting of randomly selected neurons will be impaired by the  $r_{sc}$  structure in a manner proportional to the size of the ensemble but that the  $r_{sc}$  structure can actually improve WM coding in  $r_{sc} + r_{signal}$ -optimized ensembles.

**Different Ensemble Configurations Optimize WM Coding When the  $r_{sc}$  Structure Is Intact vs. Removed.** The previous results demonstrate that accounting for an ensemble's  $r_{sc}$  structure can significantly alter estimates of its WM information content. A complementary question is whether accounting for the  $r_{sc}$  structure also alters estimates of individual neurons' contributions to an ensemble's WM coding. Are ensembles that maximize coding efficiency when the  $r_{sc}$  structure is intact composed of the same neurons that maximize coding efficiency when the  $r_{sc}$  structure is shuffled out? To answer this question, we examined the proportion of units common to both the  $r_{signal} + r_{sc}$  and  $r_{signal}$ -only ensembles for each ensemble size (Fig. 4E). The proportion is significantly less than 1 for en-

semble sizes of 2 to 50 neurons ( $P < 0.05$ , z-test of proportions, Hochberg-corrected), indicating that the ensembles generated by the two methods are not identical; the similarity within an individual session can be as low as 33%. The  $r_{signal} + r_{sc}$  and  $r_{signal}$ -only procedures also recruited units into ensembles in different sequences (Spearman's  $\rho < 1$  in all sessions, mean  $\rho = 0.713$ ;  $P < 0.05$ , Bonferroni-corrected) (Fig. S6). These results demonstrate that different subpopulations of neurons optimize WM coding when the intrinsic  $r_{sc}$  structure is present vs. when it is absent although some neurons strongly contribute to WM coding regardless of an ensemble's  $r_{sc}$  structure.

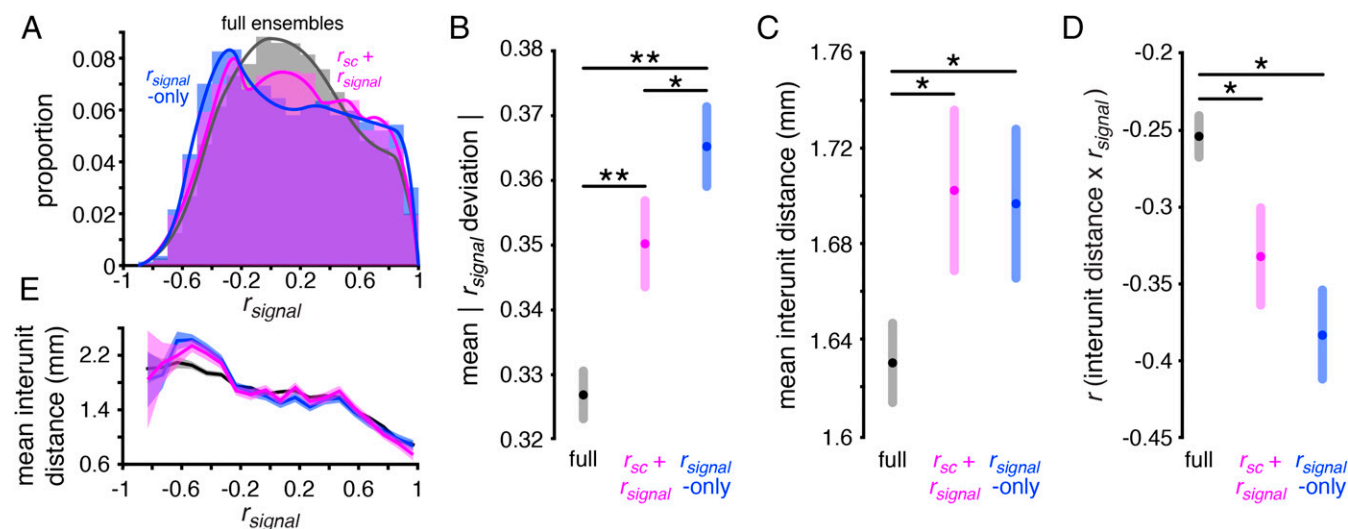
**Ensembles Optimized for WM Representation Are  $r_{signal}$ -Diverse and Anatomically Dispersed.** One of our earlier analyses demonstrated that near-maximum decoding performance can be achieved with a relatively small proportion of recorded units and that accounting for an ensemble's  $r_{signal}$  and/or  $r_{sc}$  structure can further enhance

WM coding. If the WM coding is optimized in these ensembles by maximizing their representation of the stimulus space, their  $r_{\text{signal}}$  distributions should be broader than those of the full recorded ensembles. We tested this hypothesis by examining the  $r_{\text{signal}} + r_{\text{sc}}$  and  $r_{\text{signal}}$ -only ensembles that achieved  $\geq 95\%$  of maximum decoding performance in each session (which we refer to as “near-max” ensembles). Indeed, we found that the  $r_{\text{signal}}$  distributions of the near-max  $r_{\text{signal}} + r_{\text{sc}}$  ensembles,  $r_{\text{signal}}$ -only ensembles, and full ensembles were all significantly different from each other (Fig. 5A) ( $P \ll 0.001$  for all comparisons,  $\chi^2$  test, Bonferroni-corrected). The width of the  $r_{\text{signal}}$  distribution, measured as the mean absolute deviation (*Materials and Methods*), was larger for the near-max  $r_{\text{signal}} + r_{\text{sc}}$  and  $r_{\text{signal}}$ -only ensembles than for all units (Fig. 5B) ( $P \ll 0.001$  for both,  $F$  test, Bonferroni-corrected) and larger in the  $r_{\text{signal}}$ -only than the  $r_{\text{signal}} + r_{\text{sc}}$  ensembles ( $P = 0.01$ ,  $F$  test).

Prior studies have reported weak topography for visual (44, 45) and mnemonic (29) space in LPFC; units’ tuning similarity and the anatomical distance between them—the “interunit distance”—are negatively correlated. If the optimized ensembles reflect this topography, their broader representation of the stimulus space means that they should encompass larger regions of cortex relative to the full recorded ensembles. Indeed, we found that the mean distance between units—or interunit distance—was larger in the near-max  $r_{\text{signal}} + r_{\text{sc}}$  and  $r_{\text{signal}}$ -only ensembles than the full ensembles (Fig. 5C) ( $P < 0.005$  for both,  $F$  test, Bonferroni-corrected) (*Materials and Methods*). We also found that topography in the optimized ensembles was enhanced compared with the full ensembles (Fig. 5D); the correlation between interunit distance and  $r_{\text{signal}}$  was significantly stronger in the near-max  $r_{\text{signal}} + r_{\text{sc}}$  ensembles ( $r = -0.33$ ) and  $r_{\text{signal}}$ -only ensembles ( $r = -0.38$ ) compared with the full ensembles ( $r = -0.26$ ;  $P < 0.005$  for both, bootstrap test). A potential explanation for this difference is that the distance between units with negative  $r_{\text{signal}}$  is larger in the optimized ensembles (Fig. 5E). Remarkably, the mean interunit distance can reach 2.5 mm in the near-max ensembles. Considering that cortical columns in the LPFC could span  $\sim 0.7$  mm (46),

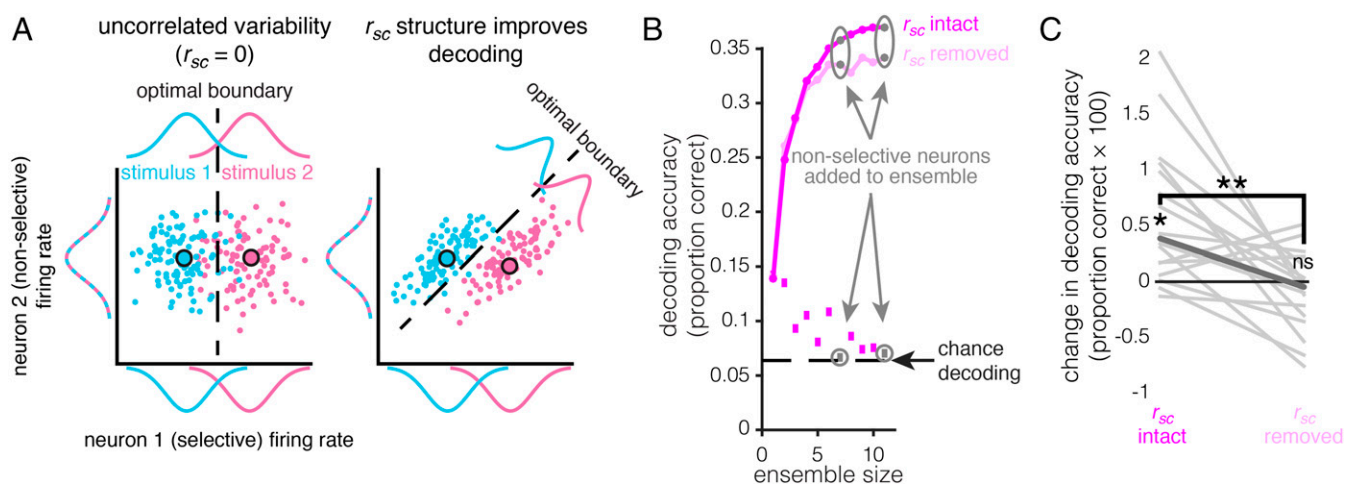
this result suggests that optimal ensembles extend across several cortical columns. These results link the spatial mnemonic topography of LPFC to principles of WM coding. They also demonstrate the utility of accounting for neuronal information content when examining cortical organization, compared with approaches that focus on neuronal tuning characteristics while leaving their effects on information implicit. These findings are also robust to the choice of near-max value because repeating the analyses with different thresholds yielded similar results (Fig. S7).

**Nonselective Units Can Improve WM Coding by Modifying the  $r_{\text{sc}}$  Structure.** Given our observation that the  $r_{\text{sc}}$  structure can significantly affect the information content of a neuronal ensemble during WM, it is possible that neurons that do not contain task-related information in isolation could still influence the information content of an ensemble by modifying the  $r_{\text{sc}}$  structure (Fig. 6A). The  $r_{\text{signal}}$  distribution of the  $r_{\text{signal}} + r_{\text{sc}}$  ensembles in Fig. 6A contains a peak near  $r_{\text{signal}} = 0$ , unlike the  $r_{\text{signal}}$ -only ensembles, suggesting that units with orthogonal and/or weak selectivity may contribute more to WM coding when the  $r_{\text{sc}}$  structure is intact. Indeed, nonselective units were sometimes added to ensembles before selective units, and before decoding performance saturated (Fig. 6B). To test whether these units were increasing ensemble WM information by modifying the  $r_{\text{sc}}$  structure, we identified all of the non delay-selective units ( $P \geq 0.05$ , one-way Kruskal–Wallis ANOVA with stimulus location as the factor) that were added before decoding performance saturation in the  $r_{\text{signal}} + r_{\text{sc}}$  ensembles (Fig. 6B) (16 units in total). We then compared the amount of information these units contributed to an ensemble before and after shuffling out the  $r_{\text{sc}}$  structure (Fig. 6C) (*Materials and Methods*). Removing the  $r_{\text{sc}}$  structure significantly decreased the amount of WM information contributed by these units ( $P < 0.01$ , signed rank test, paired), and the amount of WM information contributed after shuffling was not significantly different from zero ( $P = 0.43$ , Wilcoxon signed rank test, unpaired; additional descriptive statistics and control analyses for these units



**Fig. 5.** Ensembles optimized for WM representation are  $r_{\text{signal}}$ -diverse and anatomically dispersed. (A)  $r_{\text{signal}}$  distributions for the full ensembles (gray;  $n = 12,222$  units), near-max  $r_{\text{signal}} + r_{\text{sc}}$  ensembles (violet;  $n = 2,414$ ), and near-max  $r_{\text{signal}}$ -only ensembles (blue;  $n = 2,724$ ), pooled across all sessions. All three distributions are significantly different from each other ( $P \ll 0.001$ ,  $\chi^2$  test, Bonferroni-corrected; computed using nonoverlapping bins of size = 0.1). (B) Mean  $|r_{\text{signal}} \text{ deviation}|$  in the full (gray), near-max  $r_{\text{signal}} + r_{\text{sc}}$  (violet), and near-max  $r_{\text{signal}}$ -only ensembles (blue).  $r_{\text{signal}} \text{ deviation}$  is defined as the difference between a unit pair’s  $r_{\text{signal}}$  and the mean  $r_{\text{signal}}$  of the ensemble to which the unit pair belongs.  $**P \ll 0.001$ , Bonferroni-corrected,  $*P = 0.01$ ,  $F$  test (*SI Materials and Methods*). Shaded regions represent Bonferroni-corrected 95% comparison intervals between group means (*SI Materials and Methods*). (C) Mean interunit distance in each of the three ensemble groups.  $*P < 0.005$ ,  $F$  test, Bonferroni-corrected. Shaded regions represent Bonferroni-corrected 95% comparison intervals between group means. (D) Correlation between interunit distance and  $r_{\text{signal}}$  in the three ensemble groups.  $*P < 0.005$ , bootstrap test. Shaded regions represent bootstrapped 95% confidence intervals. (E) Mean interunit distance (y axis) as a function of  $r_{\text{signal}}$  in each of the ensemble groups, computed using nonoverlapping  $r_{\text{signal}}$  bins of size 0.1. Shaded region denotes SEM.





**Fig. 6.** Nonselective neurons can increase ensemble information by modifying the  $r_{sc}$  structure. (A) Two-neuron conceptual diagram of how a nonselective neuron could increase ensemble information content. In the first scenario (Left), one neuron differentiates between two stimuli (i.e., is selective; stimuli are denoted by blue and pink), and the other neuron does not (i.e., is not selective). The response variability of the two neurons is not correlated (i.e.,  $r_{sc} = 0$ ). In the second scenario (Right), the individual neurons' properties are identical, yet correlated response variability (i.e., the  $r_{sc}$  structure) improves discrimination between the two stimuli relative to the uncorrelated scenario. (B) The continuous line plots with circular markers show the ensemble decoding accuracy (y axis) as a function of size (x axis) for the  $r_{sc} + r_{signal}$ -optimized method for a single example ensemble, before decoding saturation, for  $r_{sc}$ -intact data (magenta) and  $r_{sc}$ -shuffled data (pale magenta). The square markers at the bottom of the plot denote the decoding accuracy (y axis) of the individual unit added to the ensemble at a given size (x axis). Notice units that are added to the population that are not selective (gray). (C) Change in decoding accuracy from adding nonselective units to presaturation ensembles (y axis) when the  $r_{sc}$  structure is intact (left) and removed (right). Each line is the change for an individual unit. The bolded line is the median. Removing the  $r_{sc}$  structure eliminates the information gain contributed by these units. \* $P = 0.001$ , signed-rank test; \*\* $P < 0.003$ , paired signed-rank test; ns (not significant),  $P = 0.43$ , signed-rank test;  $n = 16$ .

are provided in Fig. S8). We also found 15 nonselective, noise-shaping neurons during the stimulus epoch. Only one of the nonselective, noise-shaping neurons was common to both epochs. However, the decoding improvement contributed by these neurons, both before and after removing the  $r_{sc}$  structure, was more consistent during the delay than the stimulus epoch (Fig. S9). These results demonstrate the existence of nonselective noise-shaping neurons: neurons that do not contain task-related information in isolation but increase the information content of an ensemble entirely through modifying the  $r_{sc}$  structure.

## Discussion

By using microelectrode arrays to record from ensembles of LPFC neurons, we were able to elucidate the effects of the  $r_{sc}$  structure on WM coding and, more generally, how WM is represented in neuronal ensembles. We found that the relationship between  $r_{sc}$  and  $r_{signal}$  during WM maintenance was as predicted by connection topography in which similarly tuned neurons are recurrently excitatory and dissimilarly tuned neurons are mutually inhibitory. Using a linear decoder, we found that removing the  $r_{sc}$  structure could increase or decrease the information content of the neuronal ensemble, depending on the size and composition of the ensemble. Consistent with previous findings, WM fidelity in ensembles of randomly selected neurons was impaired by the  $r_{sc}$  structure, and the magnitude of the impairment was proportional to the size of the ensemble. However, the intrinsic  $r_{sc}$  structure improved WM coding in smaller neuronal ensembles optimized for WM representation ( $r_{signal} + r_{sc}$  ensembles). The  $r_{signal} + r_{sc}$  ensembles consisted of different neurons than ensembles optimized for WM representation in the absence of the  $r_{sc}$  structure ( $r_{signal}$ -only ensembles). The  $r_{signal} + r_{sc}$  ensembles had a broader  $r_{signal}$  distribution, were more anatomically dispersed, and exhibited stronger topography than the full population of recorded LPFC units. Finally, we found that individual units that did not encode WM in isolation (nonselective neurons) could still contribute to WM coding when part of an ensemble by altering the ensemble's  $r_{sc}$  structure.

**Recurrent Network Dynamics During WM Coding.** WM representations in the LPFC are hypothesized to be maintained by a network structure of recurrent excitation and lateral inhibition (12–18). The resulting dynamics should manifest as changes in  $r_{sc}$  during WM maintenance (delay epoch) relative to other epochs. We observed this phenomenon in our data—mean  $r_{sc}$  is lower during the stimulus epoch compared with the delay epoch. A previous experiment (30) reported this trend but did not find a significant effect, perhaps due to a smaller sample size (295 pairs, compared with our 10,535 pairs). A second prediction is that WM maintenance should modify  $r_{sc}$  as a function of  $r_{signal}$ ;  $r_{sc}$  should be lower between neurons with dissimilar tuning than neurons with similar tuning (18, 29–33). Indeed, we found that the relationship between  $r_{sc}$  and  $r_{signal}$  changed as predicted during the delay, compared with the fixation and stimulus epochs (Fig. 2B). Our findings indicate that WM maintenance modulates the  $r_{sc}$  structure of LPFC neuronal ensembles in a manner consistent with recurrent excitation and lateral inhibition.

**Decoding WM Representations from LPFC Neuronal Ensembles.** A prior study showed that using a pseudopopulation (asynchronously recorded neurons) of the eight most informative LPFC neurons to decode spatial WM information during a match/nonmatch WM task yielded nearly identical results as using the entire 600-neuron pseudopopulation (39). We also showed that a small subensemble of the most informative neurons contain nearly as much WM information as the full recorded population. Importantly, we demonstrated that accounting for the  $r_{sc}$  structure increases ensemble efficiency; thus, pseudopopulation analyses likely overestimate the number of neurons required to achieve a given decoding accuracy.

A second study (40) using simultaneous recordings from 32 electrodes was also able to decode which of eight locations was being remembered during a spatial WM task. However, their study was primarily concerned with how cortical depth affected the ability to decode a remembered location from local field potentials

(LFPs) and contained minimal analysis of spiking activity or the impact of neuronal ensemble composition on WM coding.

**Effects of  $r_{sc}$  and  $r_{signal}$  Structures on WM Coding.** The observed patterns of  $r_{sc}$  and  $r_{signal}$  are thought to be indicative of a network structure that stabilizes WM representations over time (18, 36, 47). Our results demonstrate that these correlations can also affect the readout of WM representations from neuronal ensembles: If WM representations are read out from optimized ensembles, then the network correlation structure will favor WM coding; however, if WM representations are read out from ensembles that are “suboptimal,” then the correlation structure could impair WM coding. Our experiment shows that these changes in ensemble information content can reach 20%. Thus, a mechanism that is thought to temporally stabilize WM representations can also affect the ability to read out these representations. Note that additional discussion on how our findings extend to larger neuronal ensembles and on the effects of spike sorting in our analyses can be found in *SI Discussion*.

**Effects of the  $r_{sc}$  Structure on Information in Non-WM Tasks.** Reports of the effects of ensemble  $r_{sc}$  structure on information content vary significantly in sign and magnitude (19, 22, 24, 27, 42, 48, 49); our results can help reconcile these disparate accounts. For example, previous studies that applied decoding techniques to simultaneously recorded ensembles found that removing the  $r_{sc}$  structure decreased decoding accuracy for grating orientation (48) and remembered location (49) whereas another study reported a positive effect of pairwise  $r_{sc}$  on information coding (32). Moreover, spatial attention increases signal-to-noise primarily by reducing  $r_{sc}$  (25–27). We found that the effect of  $r_{sc}$  structure on ensemble information varied dramatically depending on an ensemble’s size and composition; removing the  $r_{sc}$  structure from the full recorded ensembles increased decoding accuracy, but removing the  $r_{sc}$  structure from the most informative subensembles decreased decoding accuracy. The discrepancies across previous studies may arise from the location in configuration space of the neuronal ensembles under investigation (50). Importantly, they should caution us against making broad conclusions concerning how variables such as  $r_{sc}$  shape information transmission across brain areas. To fully clarify this issue one must identify which neurons are contributing to coding, which poses a significant technical challenge.

Our ensemble construction procedures were designed in part to sidestep the challenge of identifying which neurons contribute to coding and to allow us to characterize the system at specific states of interest. One may argue that we did not examine the full ensemble configuration space. Such an undertaking would be computationally infeasible; there are  $\sim 10^{15}$  unique ensembles that could be created from 50 neurons. Thus, our results may actually underestimate the true range of effect sizes. Nevertheless, even a limited search of the full configuration space demonstrates the importance of the  $r_{sc}$  structure to WM coding in LPFC neuronal ensembles.

**Noise-Shaping Units.** Neurophysiological studies typically assume that, if an ensemble codes for some behavior, the individual neurons constituting that ensemble will also code for that behavior when examined in isolation. This assumption is implicit in the method that forms the bedrock of neurophysiological research: serial recording of individual neurons. However, this approach cannot account for simultaneity between neurons. The use of large-scale simultaneous ensemble recordings allowed us to find nonselective noise-shaping neurons: neurons that are not selective for a remembered location but can improve the fidelity of WM representation in an ensemble by modifying the ensemble’s  $r_{sc}$  structure (Fig. 6A). A similar phenomenon was shown in a prior fMRI study; voxels that do not contain stimulus information in isolation can improve decoding when part of an “ensemble” of voxels (51). Their study and ours seem to report two different

instances of the same general property of information coding in multidimensional systems: Features (e.g., voxels or neurons) that do not contain information in isolation can still modify the amount of information in a system to which they belong by changing the structure of correlated variability. It remains to be observed whether nonselective noise-shaping neurons contribute to information coding in other tasks and brain regions.

## Conclusion

We leveraged the simultaneous multineuron recording capabilities of microelectrode arrays to elucidate how WM is coded in LPFC neuronal ensembles. We found that the structure of the correlated variability ( $r_{sc}$ ) supports current computational models of how sustained activity emerges in WM networks. A great deal of the power of modeling studies lies in their ability to explore parameter spaces, and we devised our ensemble construction procedures in an attempt to create an empirical analog of this capability. Applying these procedures revealed that the size,  $r_{signal}$  structure, and  $r_{sc}$  structure of an ensemble can profoundly impact WM coding. We also found that LPFC neuronal ensembles that optimize the coding of remembered locations are heterogeneously tuned and anatomically dispersed. Finally, we demonstrated that a ubiquitous assumption in neurophysiological studies—that only “selective” neurons contribute to information coding—is not justified in LPFC networks; nonselective neurons can contribute to information coding by shaping the  $r_{sc}$  structure. More generally, our results emphasize the relevance of ensemble-level phenomena in building a comprehensive understanding of brain networks.

## Materials and Methods

**Ethics Statement.** The animal care and ethics are identical to those in ref. 45, were in agreement with Canadian rules and regulations, and were pre-approved by the McGill University Animal Care Committee. Full details can be found in *SI Materials and Methods*.

**Task.** Trials were separated into four epochs: fixation, stimulus presentation (stimulus), delay, and response (Fig. 1A). The animal initiated a trial by maintaining gaze on a central fixation spot (0.08 degrees<sup>2</sup>) and pressing a lever; the subject needed to maintain fixation within 1.4° of the spot until cued to respond. The fixation period lasted either 483, 636, or 789 ms, determined randomly at the beginning of each trial. After fixation, a sine-wave grating (2.5 Hz/degree, 1° diameter, vertical orientation) appeared at one of 16 randomly selected locations for 505 ms. The potential stimulus locations were arranged in a 4 × 4 grid, spaced 4.7° apart, centered around the fixation point. The stimulus period was followed by a randomly variable delay period of 496 to 1,500 ms. The delay period ended and the response period commenced when the fixation point was extinguished, cuing the animal to make a saccade to the location of the previously presented stimulus and then to release the lever. The animal had 650 ms to respond. Successful completion of the trial yielded a juice reward. The minimum duration between trials was 300 ms. Fixation breaks during the trial or failure to saccade to the target in the allotted time resulted in immediate trial abortion without reward and a delay of 3.5 s before the next trial could be initiated.

**Experimental Setup.** The experimental setup is identical to those in refs. 27 and 45. Full details can be found in *SI Materials and Methods*.

**Microelectrode Array Implant.** As in refs. 27 and 45, we chronically implanted a 10 × 10, 1.5-mm microelectrode array (Blackrock Microsystems LLC) (52, 53) in each monkey’s left LPFC—anterior to the knee of the arcuate sulcus and caudal to the posterior end of the principal sulcus (area 8a) (Fig. 1B). Detailed surgical procedures can be found in ref. 45.

**Recordings and Spike Detection.** Data were recorded using a Cerebus Neuronal Signal Processor (Blackrock Microsystems LLC) via a Cereport adapter. Spike waveforms were detected online by thresholding. The extracted spikes (48 samples at 30 kHz) were resorted manually in OfflineSorter (Plexon Inc.). The electrodes on each MEA were separated by at least 0.4 mm and were organized into three blocks of 32 electrodes. We collected data from one block during each recording session. Detailed recording procedures can be found in ref. 45. We collected spike data across 12 recording sessions (7 in JL, 5 in F), yielding a



total of 545 units: 164 single neurons (99 in JL, 65 in F) and 381 multiunits (221 in JL, 160 in F). Multiunits were defined as threshold-crossing events, with action potential-like morphology, that were not similar enough to be included with any of the well-defined single units. Units with mean firing rates of less than 0.5 Hz during the stimulus or delay epoch and units that fired in fewer than 5% of trials were excluded from analysis.

**Analysis Epochs.** We analyzed the final 483 ms of the fixation epoch, the initial 496 ms of the stimulus epoch, and the initial 496 ms of the delay epoch. These durations were selected to make the analysis epochs as similar as possible, given the constraints that the stimulus presentation software operated at a resolution of 85 Hz, and to include as many trials as possible. We analyzed only successfully completed trials. Data analysis was performed using MATLAB.

**Spatial Selectivity.** To determine whether a unit was selective for the stimulus location during a given epoch, we computed a one-way Kruskal–Wallis ANOVA on epoch-averaged firing rates, with stimulus location as the independent variable. A unit was defined as selective if the test resulted in  $P < 0.05$ . The number of trials per location varied across recording sessions (mean = 21.5, minimum = 8, maximum = 36). Thus, the total number of trials across all locations ( $n$  trials  $\times$  16 locations) ranged from 128 to 576; 417 (76%) of the 545 recorded units exhibited delay epoch selectivity.

**Spike Count Correlation and Signal Correlation Analysis.** To compute  $r_{sc}$ , we first z-scored each unit's spike counts for each condition (i.e., stimulus location) in each epoch. Z-scoring separately for each condition removes the spike rate variability across conditions due simply to variability in firing rate responses to different stimuli (i.e., stimulus selectivity); z-scoring also removes differences in baseline firing rates for different neurons. We then grouped units into simultaneously recorded pairs ( $n = 12,006$ ) and computed Pearson's correlation coefficients ( $r_{sc}$ ) between the z-scored spike counts (21, 45). We minimized the risk of falsely inflating the correlation values by excluding correlations between units on the same electrode from analysis.

$r_{sc}$  can covary with firing rate (21, 35) so, to ensure that differences in  $r_{sc}$  across epochs were not confounded by differences in firing rates, we implemented a distribution-matching procedure as in refs. 34 and 54. To create matched distributions, we first computed the distribution of geometric mean firing rates for every pair of neurons included in the  $r_{sc}$  analysis, for each epoch. Next, we computed the greatest common distribution present across all epochs. The distributions in each epoch were matched to the common distribution by randomly discarding data points from each bin of an epoch's distribution, until each epoch's bins contained the same number of data points as those of the common distribution, which reduced the number of correlation pairs from 12,006 to 10,535. This distribution-matching procedure was repeated 2,000 times. The mean of these 2,000 distributions is plotted in Fig. 2A. We used a bootstrap test to determine whether the mean pairwise  $r_{sc}$  during a given epoch was different from zero and/or different from other epochs. We first computed the mean within each of the 2,000 firing rate-matched  $r_{sc}$  distributions, yielding 2,000 estimates of the mean. If the 0.1th percentile of this distribution of 2,000 mean  $r_{sc}$  values was greater than zero, we deemed it significantly greater than zero. If the central 99.9% of the distributions of mean  $r_{sc}$  values for two epochs did not overlap, we deemed them significantly different.

Signal correlation ( $r_{signal}$ ) was computed as the correlation of two neurons' mean responses to each of the 16 stimuli. To determine whether tuning similarity affects  $r_{sc}$ , we performed a bootstrap test similar to the one described above. For each of the 2,000 firing rate-matched  $r_{sc}$  distributions, we subdivided the  $r_{sc}$  values based on their corresponding  $r_{signal}$  value. Neuron pairs with  $r_{signal} > 0.25$  were categorized as "similarly tuned" and pairs with  $r_{signal} < -0.25$  as "dissimilarly tuned." We then computed the median  $r_{sc}$  within each of the 2,000 groups of similarly and dissimilarly tuned neuron pairs, yielding six distributions of median  $r_{sc}$  values (three epochs  $\times$  two tuning similarity groups). If the central 99.9% of two distributions of median  $r_{sc}$  values did not overlap, we deemed them significantly different.

**Population Decoding.** We used a support vector machine (SVM) (Libsvm 3.14), a linear classifier, to extract task-related activity from the population-level representations of simultaneously recorded neuronal ensembles (55, 56). The SVM used epoch-averaged firing rate data from an ensemble to predict at which of the 16 locations the stimulus appeared in a given trial during each the stimulus and delay epochs. Each neuron constitutes a dimension in the multidimensional population space, and the SVM seeks to find the boundaries that best distinguish between population responses to each stimulus location. Units that fired in fewer than 5% of trials or fired at a mean rate of less than 0.5 Hz were excluded from analysis. We scaled each unit's firing

rates to  $[-1, 1]$  by subtracting the midrange rate value ( $\max + \min$ )/2 and dividing by one-half the range ( $\max - \min$ )/2, to prevent units with larger absolute changes in firing rate from dominating the classification boundaries. These two parameters were determined from the training set and applied to both the training and testing sets. We assessed the classifier's performance using fivefold cross-validation, such that 80% of the trials were used to train the decoder (the "training set"), and the decoder attempts to classify the remaining 20% of the trials (the "testing set").

To test whether the  $r_{sc}$  structure affects the fidelity of ensemble representations during WM, we removed the  $r_{sc}$  structure from the neural data using a shuffling procedure identical to that described in ref. 27. The shuffling procedure consisted of randomizing the trial order within each location condition for each neuron, such that the condition (i.e., the remembered location) for a given trial remained the same for all neurons, but the firing rates for each neuron were drawn from different trials. This procedure destroys the simultaneity, and thus the intrinsic  $r_{sc}$  structure, in the recordings. The decoding analysis was then run on the shuffled firing rates. The shuffled decoding procedure was repeated 200 times, and the mean of these 200 iterations was taken as the " $r_{sc}$ -shuffled" decoding accuracy. We refer to the percent change in decoding accuracy due to shuffling as  $\Delta_{shuffled}$ , defined as  $[(accuracy_{rsc-shuffled}/accuracy_{rsc-intact}) - 1] \times 100\%$ .

We define coding efficiency as the amount of WM information in an ensemble of a given size relative to another ensemble of the same size, computed as  $[(accuracy_{ensemble1}/accuracy_{ensemble2}) - 1] \times 100\%$ .

To ensure that our results were robust to the choice of classifier, we repeated our decoding analyses using logistic regression instead of SVM. The data analysis procedure for the logistic regression was identical to that of SVM (i.e., excluding low firing rate units, scaling firing rates, and cross-validation procedure), except we used the LIBLINEAR library (57) to perform logistic regression instead of SVM. As in the SVM, each unit is a predictor, there are no interaction terms, and the neuronal firing rates are used to predict the remembered location. In the context of our analysis, the relevant differences between SVM and logistic regression are that they have different loss (sometimes called error) functions and that logistic regression is probabilistic whereas SVM is deterministic (see ref. 57 for more detail).

**Functional Anatomy.** To compare the widths of the  $r_{signal}$  distributions of the near-max  $r_{sc} + r_{signal}$  ensembles, near-max  $r_{signal}$ -only ensembles, and full recorded ensembles, we first computed the absolute value of the  $r_{signal}$  deviation ( $|r_{signal}|$  deviation) of each unit pair in an ensemble. The  $r_{signal}$  deviation is defined as the difference between a unit pair's  $r_{signal}$  and the mean  $r_{signal}$  of the ensemble to which the unit pair belongs. We then assessed the difference between the  $|r_{signal}|$  deviations in each type of ensemble ( $r_{sc} + r_{signal}$ ,  $r_{signal}$ -only, and full recorded ensemble) by fitting a linear mixed-effects model with ensemble type as a fixed effect and recording session as a random effect to predict  $|r_{signal}|$  deviation. Pairwise significance between ensemble types was determined by a Bonferroni-corrected  $F$  test on the difference between the two groups' coefficients, the degrees of freedom approximated using the Satterthwaite equation. Distance between units (interunit distance) in each ensemble type was compared using a similar linear mixed-effects model, but to predict interunit distance. The measures of variability displayed in Fig. 6B and C are Bonferroni-corrected simultaneous comparison intervals, generated using equation 3.32 in ref. 58.

The strength of topography was assessed by computing the Pearson correlation between every simultaneously recorded unit pair's  $r_{signal}$  and interunit distance within each group of ensembles ( $r_{sc} + r_{signal}$ ,  $r_{signal}$ -only, and full recorded ensemble). Significance was assessed using a bootstrap test: The distributions of  $r_{signal}$  and interunit distance were randomly sampled, with replacement, to generate new vectors of length equal to the original distributions, and the Pearson correlation was computed between the new resampled vectors. This procedure was repeated 10,000 times to generate a distribution of correlation coefficients. The strength of topography for two ensemble groups was significantly different at  $P < \alpha$  if the central proportion of size  $1 - \alpha$  of the two groups' bootstrapped correlation coefficients did not overlap.

**Nonselective Noise-Shaping Neuron Analysis.** To find nonselective noise-shaping neurons, we examined the results of the  $r_{signal} + r_{sc}$ -optimized procedure and identified all instances in which adding a non-delay-selective neuron ( $P \geq 0.05$ , one-way Kruskal–Wallis ANOVA, firing rate  $\times$  location) increased the information content of the ensemble. We then used the near-max  $r_{signal} + r_{sc}$  ensembles to decode firing rate data from which the  $r_{sc}$  structure had been shuffled out and compared the amount of information these units contributed to an ensemble before and after shuffling. Note that using the near-max  $r_{signal} + r_{sc}$  ensembles to decode firing data from which the  $r_{sc}$  structure has been shuffled out is not the same as the  $r_{signal}$ -only optimized procedure. Although both involve decoding  $r_{sc}$ -shuffled firing rates, in the

former case, the ensembles are generated using  $r_{sc}$ -intact data whereas, in the latter case, the ensembles are generated using  $r_{sc}$ -shuffled data.

We performed a control analysis to determine whether WM-selective units that contribute similar amounts of information to an ensemble as nonselective noise-shaping units also contribute information by modifying the  $r_{sc}$  structure (Fig. S8). We accomplished this analysis by using a distribution-matching procedure similar to the one used to match distributions of firing rates in the pairwise  $r_{sc}$  analysis. First, we computed the distributions of improvements in decoding accuracy from nonselective noise-shaping units and selective units. Next, we randomly discarded data points from each bin of the selective units' distribution until it matched the nonselective noise-shaping units' distribution. The same data points were discarded from the distribution of selective units' decoding accuracy on the  $r_{sc}$ -shuffled firing

rates. This procedure was repeated 2,000 times. We then computed the median within each of these 2,000 matched distributions for  $r_{sc}$ -intact data and the  $r_{sc}$ -shuffled data. If the central 95% of the two distributions of median improvement in decoding accuracy did not overlap, we deemed them significantly different. We also performed an additional control analysis using a standard ANOVA to assess selectivity instead of a nonparametric Kruskal–Wallis ANOVA.

**ACKNOWLEDGMENTS.** We thank Walter Kucharski for fabrication expertise; Stephen Nuara for zoological talents; Megan Schneiderman for assistance with recordings; Rishi Rajalingham for scrutiny and code; Dr. Ruben Moreno-Bote, Ramon Nogueira, Roberto Gulli, and Lyndon Duong for helpful discussion; and the J.C.M.-T. laboratory for support and funding.

- Baddeley AD, Hitch G (1974) Working memory. *The Psychology of Learning and Motivation: Advances in Research and Theory*, ed Bower GH (Academic, New York), Vol 8.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
- Hebb DO (2005) *The Organization of Behavior: A Neuropsychological Theory* (John Wiley & Sons, New York).
- Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. *Science* 173(3997):652–654.
- Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol* 61(2):331–349.
- Batuev AS (1986) Neuronal mechanisms of goal-directed behavior in monkeys. *Neurosci Behav Physiol* 16(6):459–465.
- Gnadt JW, Andersen RA (1988) Memory related motor planning activity in posterior parietal cortex of macaque. *Exp Brain Res* 70(1):216–220.
- Mendoza-Halliday D, Torres S, Martinez-Trujillo JC (2014) Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat Neurosci* 17(9):1255–1262.
- Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16(16):5154–5167.
- Riley MR, Constantinidis C (2016) Role of prefrontal persistent activity in working memory. *Front Syst Neurosci* 9:181.
- Quiari Quiroga R, Panzeri S (2009) Extracting information from neuronal populations: Information theory and decoding approaches. *Nat Rev Neurosci* 10(3):173–185.
- Wang X-J (2001) Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci* 24(8):455–463.
- Constantinidis C, Wang X-J (2004) A neural circuit basis for spatial working memory. *Neuroscientist* 10(6):553–565.
- Durstewitz D, Seamans JK, Sejnowski TJ (2000) Neurocomputational models of working memory. *Nat Neurosci* 3(Suppl):1184–1191.
- Compte A, Brunel N, Goldman-Rakic PS, Wang X-J (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb Cortex* 10(9):910–923.
- Amit DJ, Brunel N (1997) Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb Cortex* 7(3):237–252.
- Camperi M, Wang XJ (1998) A model of visuospatial working memory in prefrontal cortex: Recurrent network and cellular bistability. *J Comput Neurosci* 5(4):383–405.
- Wimmer K, Nykamp DQ, Constantinidis C, Compte A (2014) Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat Neurosci* 17(3):431–439.
- Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7(5):358–366.
- Moreno-Bote R, et al. (2014) Information-limiting correlations. *Nat Neurosci* 17(10):1410–1417.
- Cohen MR, Kohn A (2011) Measuring and interpreting neuronal correlations. *Nat Neurosci* 14(7):811–819.
- Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370(6485):140–143.
- Shadlen MN, Newsome WT (1994) Noise, neural codes and cortical organization. *Curr Opin Neurobiol* 4(4):569–579.
- Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11(1):91–101.
- Cohen MR, Maunsell JHR (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12(12):1594–1600.
- Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63(6):879–888.
- Tremblay S, Pieper F, Sachs A, Martinez-Trujillo J (2015) Attentional filtering of visual information by neuronal ensembles in the primate lateral prefrontal cortex. *Neuron* 85(1):202–215.
- Romo R, Hernández A, Zainos A, Salinas E (2003) Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron* 38(4):649–657.
- Constantinidis C, Franowicz MN, Goldman-Rakic PS (2001) Coding specificity in cortical microcircuits: A multiple-electrode analysis of primate prefrontal cortex. *J Neurosci* 21(10):3646–3655.
- Constantinidis C, Goldman-Rakic PS (2002) Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *J Neurophysiol* 88(6):3487–3497.
- Qi X-L, Constantinidis C (2012) Correlated discharges in the primate prefrontal cortex before and after working memory training. *Eur J Neurosci* 36(11):3538–3548.
- Markowitz DA, Curtis CE, Pesaran B (2015) Multiple component networks support working memory in prefrontal cortex. *Proc Natl Acad Sci USA* 112(35):11084–11089.
- Katsuki F, et al. (2014) Differences in intrinsic functional organization between dorsolateral prefrontal and posterior parietal cortex. *Cereb Cortex* 24(9):2334–2349.
- Ruff DA, Cohen MR (2014) Attention can either increase or decrease spike count correlations in visual cortex. *Nat Neurosci* 17(11):1591–1597.
- de la Rocha J, Doiron B, Shea-Brown E, Josić K, Reyes A (2007) Correlation between neural spike trains increases with firing rate. *Nature* 448(7155):802–806.
- Laing CR, Troy WC, Gutkin B, Ermentrout GB (2002) Multiple bumps in a neuronal model of working memory. *SIAM J Appl Math* 63:62–97.
- Kanitscheider I, Coen-Cagli R, Kohn A, Pouget A (2015) Measuring Fisher information accurately in correlated neural populations. *PLOS Comput Biol* 11(6):e1004218.
- Rigotti M, et al. (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497(7451):585–590.
- Meyers EM, Qi X-L, Constantinidis C (2012) Incorporation of new information into prefrontal cortical activity after learning working memory tasks. *Proc Natl Acad Sci USA* 109(12):4651–4656.
- Markowitz DA, Wong YT, Gray CM, Pesaran B (2011) Optimizing the decoding of movement goals from local field potentials in macaque cortex. *J Neurosci* 31(50):18412–18422.
- Averbeck BB, Lee D (2006) Effects of noise correlations on information encoding and decoding. *J Neurophysiol* 95(6):3633–3644.
- Shamir M, Sompolinsky H (2006) Implications of neuronal diversity on population coding. *Neural Comput* 18(8):1951–1986.
- Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3:1157–1182.
- Suzuki H, Azuma M (1983) Topographic studies on visual neurons in the dorsolateral prefrontal cortex of the monkey. *Exp Brain Res* 53(1):47–58.
- Leavitt ML, Pieper F, Sachs A, Joobler R, Martinez-Trujillo JC (2013) Structure of spike count correlations reveals functional interactions between neurons in dorsolateral prefrontal cortex area 8a of behaving primates. *PLoS One* 8(4):e61503.
- Bugbee NM, Goldman-Rakic PS (1983) Columnar organization of corticocortical projections in squirrel and rhesus monkeys: Similarity of column width in species differing in cortical volume. *J Comp Neurol* 220(3):355–364.
- Polk A, Litwin-Kumar A, Doiron B (2012) Correlated neural variability in persistent state networks. *Proc Natl Acad Sci USA* 109(16):6295–6300.
- Graf ABA, Kohn A, Jazayeri M, Movshon JA (2011) Decoding the activity of neuronal populations in macaque primary visual cortex. *Nat Neurosci* 14(2):239–245.
- Graf ABA, Andersen RA (2015) Predicting oculomotor behaviour from correlated populations of posterior parietal neurons. *Nat Commun* 6:6024.
- Hu Y, Zylberberg J, Shea-Brown E (2014) The sign rule and beyond: Boundary effects, flexibility, and noise correlations in neural population codes. *PLOS Comput Biol* 10(2):e1003469.
- Yamashita O, Sato M-A, Yoshioka T, Tong F, Kamitani Y (2008) Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. *Neuroimage* 42(4):1414–1429.
- Maynard EM, Nordhausen CT, Normann RA (1997) The Utah intracortical electrode array: A recording structure for potential brain-computer interfaces. *Electroencephalogr Clin Neurophysiol* 102(3):228–239.
- Normann RA, Maynard EM, Rousche PJ, Warren DJ (1999) A neural interface for a cortical vision prosthesis. *Vision Res* 39(15):2577–2587.
- Churchland MM, et al. (2010) Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nat Neurosci* 13(3):369–378.
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273.
- Chang CC, Lin CJ (2011) LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol* 2(3):27.
- Fan R-E, Chang K-W, Hsieh C-J, Wang X-R, Lin C-J (2008) LIBLINEAR: A library for large linear classification. *J Mach Learn Res* 9:1871–1874.
- Hochberg Y, Tamhane AC (1987) *Multiple Comparison Procedures* (John Wiley & Sons, Hoboken, NJ), 1st Ed, pp 72–109.