Title
The neural architecture of working memory: anatomical and functional studies of prefrontal cortex

Author
Miller, Jacob

Publication Date
2021

Peer reviewed|Thesis/dissertation

The neural architecture of working memory:

anatomical and functional studies of prefrontal cortex

by

Jacob Adam Miller

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Neuroscience

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mark D'Esposito, Chair
Professor Richard B. Ivry
Professor Michael Silver
Professor Kevin S. Weiner

Fall 2021

Abstract

Anatomical and functional substrates for prefrontal cortex control of working memory

by

Jacob Adam Miller

Doctor of Philosophy in Neuroscience

University of California, Berkeley

Professor Mark D'Esposito, Chair

"Memory is a gift of nature, the ability of living organisms to retain and to utilize acquired information or knowledge… Owners of biological memory systems are capable of behaving more appropriately at a later time because of their experiences at an earlier time, a feat not possible for organisms without memory."

- Tulving (1995, *p. 751*)

"Let me begin with the perplexity. Man's frontal lobes have always presented problems that seemed to exceed those encountered in studying other regions of his brain."

- Teuber (1964)

In the first chapter, I investigate the role of prefrontal cortex activity in working memory. Here, I am motivated by inconsistencies in the neural substrates for working memory across studies, species, and recording techniques. For instance, non-human primate electrophysiology research finds that prefrontal circuitry maintains working memory representations, while human neuroimaging suggests that working memory content is instead stored in sensory cortices. These seemingly incompatible accounts for working memory are often confounded by differences in the amount of task training and stimulus exposure across studies, suggesting that long-term learning may influence the role of prefrontal function in working memory maintenance. To answer these questions, we longitudinally trained and scanned participants on a working memory task with complex stimuli. Then, we used multivariate analyses of functional neuroimaging (fMRI) data to test how representational structures of working memory activity patterns in prefrontal cortex change across intensive learning. We show that human prefrontal cortex develops stimulus-selective working memory responses with learning, more akin to results from electrophysiology studies. This approach uses a unique training and analysis framework to establish novel evidence for long-term memory influences on working memory maintenance.

In the second chapter, we investigate how working memory is constantly used to guide our moment to moment behaviors. This reliance on working memory can lead us to make mistakes, like saying aloud the wrong word in a conversation. Such "action slips" are common occurrences but especially pronounced in individuals with prefrontal lesions, who may often pour salt instead of sugar into one's coffee, or mistakenly type "pizza" in an immediate texting conversation when thinking about your upcoming lunch (Lhermitte et al. 1986). To study this interaction between working memory and ongoing behavior, I implemented a dual-task experiment in which directional words must be held in working memory while more immediate, but unrelated, motor movements are performed. We show that motor behaviors unrelated to current working memory information are still influenced by one's working memory content (Miller et al. 2020). We are currently testing these behaviors with predictions from cortico-striatal circuit models of working memory gating (e.g., O'Reilly and Frank, 2006) by using transcranial magnetic stimulation. By causally perturbing prefrontal functioning and cortico-striatal connectivity, can we alter when and how often working memory content influences our immediate actions?

In the final chapter, I outline how investigating human-specific neuroanatomical structures in frontal cortex is critical for a wider investigation of human cognition. The prefrontal cortex is disproportionately expanded in the human brain even relative to other advanced primates, and some structures such as tertiary sulci, small folds in the cerebral cortex, are often human-specific. I use multi-modal neuroimaging data to investigate relationships between microstructural and functional properties in human prefrontal cortex. We show that careful identification of often overlooked individual-level anatomical features (such as tertiary sulci) serve as a bridge between the microanatomical and functional properties of prefrontal cortex (Miller et al. 2021). Identifying these structures has implications for both individual-level prefrontal functioning and broader mappings between prefrontal anatomy, functioning, and cognitive domains. We propose that such careful investigations of individual-level neuroanatomy will help to generate structural-functional relationships in areas of cortex previously thought to have little or no consistent links between individual-level structure and function (Miller et al. 2021).

Table of Contents

Acknowledgements

"Life is a good teacher and a good friend. Things are always in transition, if we could only realize it. Nothing ever sums itself up in the way that we like to dream about. The off-center, in-between state is an ideal situation, a situation in which we don't get caught and we can open our hearts and minds beyond limit."

- Pema Chödrön, *When Things Fall Apart*

There are too many people to name here who have helped me along this path, both before and during my time at Berkeley. I list some of them here, but to everyone else who has inspired me scientifically and otherwise: thank you for helping me do the kind of science I never imagined. And of course, for helping me to become a more thoughtful, mature, and better person.

To the entire D'Esposito lab, starting with Mark, who has been a constant source of support. You have taught me how to embody a great balance of scientific skepticism and trust, and you have fostered a unique combination of a wildly productive, fun, and collaborative lab environment. I am especially indebted to Anastasia Kiyonaga and Arielle Tambini, as we embarked on the crazy difficult and rewarding journey of a four month long, self-experiment with 24 trips to the fMRI scanner each. Throughout that time and beyond, you have tried your best to help me be a better writer, scientist, and refine my quite poor pun delivery. To Anastasia especially, you have been like a second mentor to me, and I know your lab and trainees will flourish in your new home. I'd also like to thank postdocs Ian Ballard, Regina Lapate, and Jason Scimeca for constant support, laughter, and teaching. And to my fellow grad students inside and out of the Despolab: Elizabeth Lorenc, Dan Lurie, Adam Eichenbaum, Justin Riddle, and Holly Gildea, Zuzanna Balewski, Celia Ford, and others thank you for everything!

To Kevin Weiner, thank you for sparking and fostering my interest in, and now love for, human neuroanatomy. Many sulci await prying eyes! And to Willa Voorhies, Ben Parker, and Leana King, thank you and I know you will all thrive with Kevin. The best times in your lab are ahead of you, and I will still be Zooming in to relay jokes and annoying scientific questions.

To Michelle, who taught me the strength in not being perfect, the mindfulness in the smallest moments, and the joy of noticing when we feel worried.

And most of all, to my partner Delaney King, and friends David, Sam, Caitlin, Justin, Adam, and Sara… your love, kindness, snark, begrudging acceptance of my puns, and occasional laughs have made this difficult journey fun, fruitful, tolerable, and filled with joy. During my most anxious moments, happiest times, and frustrating paper rejections, you have been there. Thank you, thank you, thank you!

Chapter 1: Introduction

Understanding how anatomical structures of the brain support functional networks underlying human-specific aspects of cognition is a major goal in cognitive neuroscience. Of the many anatomical structures to study, the prefrontal cortex is particularly important given its central role in cognitive control and goal-directed behavior (Stuss & Knight, 2013; Miller & Cohen, 2001). Patients with prefrontal damage have deficits ranging from working memory and attention problems, to issues with motivation, response inhibition, and language. How are such a vast array of cognitive processes orchestrated by a few cubic centimeters of cortical tissue? Tackling such questions seems daunting, but I am inspired in my own work by a seminal proposal from Marsel Mesulam's "From sensation to cognition" (*Brain*, 1998). There, Mesulam lays out a theoretical framework that the unique anatomical and functional properties of association cortex circuits, such as prefrontal cortex, can account for the remarkable flexibility of human cognition. As the tools of cognitive neuroscience advance, we are now more suited than ever to carry out the empirical work to test Mesulam's framework.

With these questions in mind, I study the neuroanatomical and functional basis for prefrontal functioning with a specific interest in working memory, a core building block for flexible cognition. Maintaining and manipulating information that is no longer available to the senses underlies simple and complex behaviors that are critical for survival. Limits on the quantity and quality of representations in working memory are a primary constraint on cognition and adaptive behavior. Therefore, accurately characterizing the neural basis for working memory storage and control is paramount to understanding cognitive success and its failure.

The present work intentionally focuses on properties of the prefrontal cortex and behaviors supported by this area. However, despite self-interest, I do not aim to provide a frontal "superiority" view of brain function. Only around 80 years ago frontal cortex was viewed as having an almost "silent" or "uncommitted" contribution to human behavior (Penfield and Evans 1935). Rather, I seek to demonstrate a neuroscientific approach extendable to a broader understanding of brain function, both across scales and species. Just as the prefrontal cortex does not operate in isolation as a "seed" of cognition or conscious operations, no single study of animal neurophysiology or human neuroimaging can serve to explain the deepest questions of brain-behavior links. Instead, I firmly contend that in order to understand what processes underlie human cognition we must directly link systems-level data from animal models and the more complex neural and behavioral circuitry in humans. To that end, I hope these studies step toward the goal of determining how neuroanatomy and neural functioning give rise to behavior.

While the following chapters may seem disparate in method, they are all threaded on this central question of linking human cognition and prefrontal cortex functioning. Such spread of methods is a deliberate attempt to pull multiple lines of investigation together: studying uniquely human neuroanatomy, neural functioning, and behavior are all of the utmost importance. Emblemized by the breadth of remarkable scientists like Patricia Goldman-Rakic, we cannot let newer methods and specialization take away from our potential contributions to theoretical understanding. To best make these links, we need interpretable, theoretical frameworks that help build a convergent understanding of diverse prefrontal functioning across scales, methods, and species. In other words, working toward an "Interactionist" model of neuroscience (Badre et al. 2015).

Chapter 2: Long-term learning transforms prefrontal activity during working memory

"The significance of working memory for higher cortical function is not necessarily self-evident. Perhaps even the quality of its transient nature misleads us into thinking it is somehow less important than the more permanent archival nature of long-term memory. However, the brain's working memory function, i.e., the ability to bring to mind events in the absence of direct stimulation, may be its inherently most flexible mechanism and its evolutionarily most significant achievement."

- Goldman-Rakic (1995, *p. 483*)

This chapter contains material being prepared for submission with the following co-authors:

Arielle Tambini, Anastasia Kiyonaga, & Mark D'Esposito

Abstract

The lateral prefrontal cortex (lPFC) is reliably active during working memory (WM) across human and animal models, but there is ongoing debate regarding the role of the lPFC in successful WM. For instance, non-human primate (NHP) electrophysiology research finds that lPFC circuitry stores WM representations, while human neuroimaging suggests that WM content is instead stored in sensory cortices. These seemingly incompatible accounts for WM are often confounded by differences in the amount of task training and stimulus exposure across studies, suggesting that long-term learning may influence the role of lPFC function in WM maintenance. To test whether long-term learning influences WM representations in lPFC, we implemented a longitudinal functional MRI (fMRI) protocol in three human participants. Across three months, each participant was trained on (1) a serial reaction time (SRT) task, wherein complex fractal stimuli were embedded within probabilistic sequences, and (2) a delayed recognition task probing WM for trained or novel stimuli. Participants were scanned repeatedly across training, which allowed us to track how WM activity patterns and representations are shaped by long-term associative learning. Participants showed learning benefits in the WM task for trained, but not novel, fractals. Neurally, a significant population of voxels increased in delay activity throughout lPFC. Pattern similarity analyses of WM delay activity demonstrate that item-level representations emerged within lPFC, but not in sensory cortices, across learning. Single-item WM delay period activity patterns in lPFC also reflected sequence relationships from the SRT task, even though that information was task-irrelevant for WM. These findings demonstrate that human lPFC develops stimulus-selective WM responses with learning and suggests that long-term memory influences on WM may reconcile competing accounts of lPFC function.

Introduction

The lateral prefrontal cortex (lPFC) is considered critical for working memory (WM) across human and animal models (Funahashi et al., 1989; Goldman-Rakic, 1995; Leavitt et al., 2017; E. K. Miller et al., 2018; Sreenivasan et al., 2014). However, there is ongoing debate regarding the specific role that lPFC activity plays in successful WM (Christophel et al., 2017; Curtis & Sprague, 2021; Lara & Wallis, 2015; Mackey et al., 2016). Non-human primate (NHP) electrophysiology research typically finds that lPFC maintains feature-specific WM content (Constantinidis et al., 2018; Funahashi et al., 1989; Fuster & Alexander, 1971; Goldman-Rakic, 1995; E. K. Miller et al., 2018; Romo et al., 1999). Human neuroimaging suggests lPFC activity serves control functions over WM while feature-specific content is stored in sensory cortices instead (D'Esposito & Postle, 2015; Eriksson et al., 2015; Harrison & Tong, 2009; Riggall & Postle, 2012; Serences, 2016). However, these seemingly incompatible accounts are confounded by differences in species, measurement granularity, and the amount of task training typically performed across studies.

One possibility is that different indices of neural activity, across measurement scales, may support distinct conclusions about the cortical substrates for WM. That is, NHP studies typically record finer resolution single-unit neuronal activity compared to the millimeter scale of Blood Oxygen Level Dependent functional MRI (BOLD fMRI) (Mukamel et al., 2005; Park et al., 2017). Discrepancies between study findings may emerge if stimulus-specific WM content is represented in human lPFC via spiking patterns in single-units or populations that are simply too fine-grained for BOLD fMRI to detect, whereas the spread and spatial organization of neuronal activity in sensory cortex is more detectable at the scale of fMRI (Leavitt et al., 2017; Lorenc & Sreenivasan, 2021; Mendoza-Halliday et al., 2014; Serences, 2016). However, in some cases, stimulus-specific WM delay activity has been detected in human frontal cortex (Ester et al., 2015; Lee et al., 2013) or NHP sensory regions (Mendoza-Halliday et al., 2014; Supèr et al., 2001), highlighting the need to identify which factors truly drive observed differences in findings across studies.

In addition to differences in recording techniques between human and NHP studies, NHPs typically undergo months of training and perform orders of magnitude more task trials before the critical neural recordings occur (Berger et al., 2018; Birman & Gardner, 2016; Sarma et al., 2016). Humans typically complete only a few minutes of task practice prior to fMRI scanning. Differences observed in neural WM substrates across species may therefore be driven by long-term learning influences from extensive task and stimulus experience. In fact, the few studies that recorded from NHPs before and after WM training found plasticity in the form of increases in the magnitude of WM delay activity and in the strength of item-level stimulus representations in anterior lPFC (Dang et al., 2021; Meyer et al., 2011; Qi et al., 2019; Riley et al., 2018; Sarma et al., 2016). This suggests that human lPFC may represent item-level information in WM, depending on the level of prior training. However, the typical timeline of fMRI research has limited our ability to understand if WM representations change with long-term learning and directly test this hypothesis.

The brain regions and neural mechanisms for WM are classically considered separate from long-term memory (LTM) systems (Squire & Zola-Morgan, 1991; Warrington & Shallice, 1969; Wickelgren, 1969). However, some WM theories predict that learned associations or semantic

links between items should be reflected during WM maintenance (LaRocque et al., 2014; Oberauer, 2009), and growing evidence suggests a common neural machinery between WM and LTM (Beukers et al., 2021; Fukuda & Woodman, 2017; Hoskin et al., 2019; Lewis-Peacock & Norman, 2014; Nee & Jonides, 2011; Ranganath et al., 2003; Ranganath & Blumenfeld, 2005). In some cases WM capacity is also greater for stimuli with extensive exposure and familiarity (Asp et al., 2021; Brady et al., 2016; Jackson & Raymond, 2008; Xie & Zhang, 2017), suggesting that WM and supporting neural mechanisms may change with stimulus experience.

Here, we examined the possibility that long-term learning transforms human lPFC WM activity. We asked whether stimulus selectivity emerges in human lPFC as a function of training, akin to the stimulus-specific WM activity patterns typically found in NHP studies. To do so, three human participants completed over 20 sessions of whole-brain fMRI along with extensive at-home training across three months. During this time, participants were continually trained on a delayed recognition WM task and a sequence learning task, which both employed a set of 18 novel fractal stimuli that were unique to each participant. First, we asked whether lPFC delay period WM activity changed in magnitude across learning. Widespread decreases in lPFC activity could suggest more automatic task processing with training. Activity increases, however, could suggest greater selectivity for the repeated task structure or individual WM stimuli, as persistent activity in WM is associated with stimulus-selective patterns (Constantinidis et al., 2018; Curtis & Sprague, 2021; Murray et al., 2017). We then tested whether representations of individual stimuli or stimulus categories emerged in multivariate WM activity patterns over the course of learning. If item-level lPFC activity patterns develop over time, that would suggest that difference in participant training may explain discrepant accounts of the role of lPFC as either a source of control over WM (from human studies) versus the storage site for WM content (from single-unit NHP studies). Alternatively, long-term learning may enhance sensory cortex representations of WM content but induce no changes in lPFC, suggesting that differences in lPFC vs sensory-based WM storage models are driven by other factors than long-term learning. Finally, to understand how WM representations are shaped by associative learning, we asked if representations of associations between stimuli in shared temporal sequences (learned outside of the WM task) were reflected within WM activity patterns. To preview the results, long-term learning changed the distribution and selectivity of lPFC WM delay activity, indicating that WM maintenance mechanisms may be flexible to the extent and nature of prior experience with the WM information. These results suggest that differences in the extent of training across species may masquerade as differences in lPFC function.

Results

Intensive training improves WM performance for trained, but not novel, stimuli

To determine how long-term learning influences cortical activity patterns underlying WM maintenance, we trained three human participants on a set of fractal stimuli that was unique to each participant (**Figure 1a**) over three months. These stimuli had no pre-existing meaning and have been used to characterize the influence of long-term associative learning on neural selectivity (Ghazizadeh et al., 2018; Kim et al., 2015; Sakai & Miyashita, 1991). These complex stimuli were chosen to encourage gradual learning and to necessitate a detailed item representation to perform the task well. During the three month study period, each participant completed approximately 24 scanning (fMRI) sessions along with at-home behavioral training sessions multiple times per week (**Figure 1b**). Here we only analyze the first 17 fMRI sessions (~13 weeks), after which point new fractals were added into the stimulus set for a second phase of the study. During each fMRI session,participants performed two primary tasks, a serial reaction time (SRT) task followed by a WM task (**Figure 1c-d**). The WM task entailed a single-item delayed recognition test wherein the WM sample was either a fractal stimulus from the training set or a novel fractal that appeared only during that session. Before the study began, participants completed one block (24 trials) of WM task practice with pilot stimuli that never appeared in the main experiment. The first time each participant saw their unique set of 18 training stimuli was during the first scanning session.The SRT task used the same 18 trained fractal stimuli, for which 12 of the stimuli were embedded in high probability sequences (**Figure 1c**). The sequences were not directly related to the goals of the WM task (which was always to remember a single item), but we took advantage of the sequence structure to analyze whether item-level WM representations reflected associations (sequence-level and categorical) present in the SRT task.

Across the course of training, behavior in the WM task improved for trained stimuli, but not for novel stimuli (**Figure 1e**). Mean WM probe accuracy (% correct) for trials with trained stimuli improved by 23% across the 17 sessions, whereas accuracy increased by 4% for novel stimuli. To characterize the change in WM performance over time, we used mixed nonlinear models with session number (1 → 17; mean-centered) and stimulus category (*trained* vs. *novel*) as predictors, and WM probe accuracy as the outcome variable, focusing on the linear term (b, see **Methods**: *Statistical methods*). There was a significant interaction between session number and stimulus category for the linear term in the model, $b{\cdot}category = 0.01$, $t(94) = 2.95$, $p = 0.004$. This interaction was driven by an increase in accuracy for trained stimuli over time (main effect of session number: $b = 0.01$, $t(46) = 3.86$, $p < 0.001$), with no reliable change for novel stimuli ($b = 0.003$, $p = 0.35$).

A complementary pattern emerged when modeling WM probe response time (RT). For RT, there was also a significant interaction between session number and stimulus category ($b{\cdot}$category = -22.5, $t(4805) = -8.01$, $p < 0.001$), and this was driven by faster responses for trained stimuli over time (main effect of session number: $b = -15.9$, $t(3612) = -2.84$, $p = 0.005$), with no reliable change for novel stimuli ($b = -6.7$, $p = 0.27$). When considering WM task accuracy and response time for the trained stimuli, nonlinear mixed models slightly outperformed linear models (*accuracy*: nonlinear Akaike information criterion [AIC] = -143.5, linear AIC = -127.5; *response time*: nonlinear AIC = 53857, linear AIC = 53906). The subsequent analyses focus on nonlinear

models, because they allow for changes to occur at different rates across the 17 sessions, but all results generalize to a linear framework.

In parallel to the WM task, participants also learned associations between individual stimuli as part of regularly occurring sequences in the SRT task. Reliable associative learning across training was evidenced by reduced response times for intact sequences in the SRT task for all participants (**SI Figure 1**).

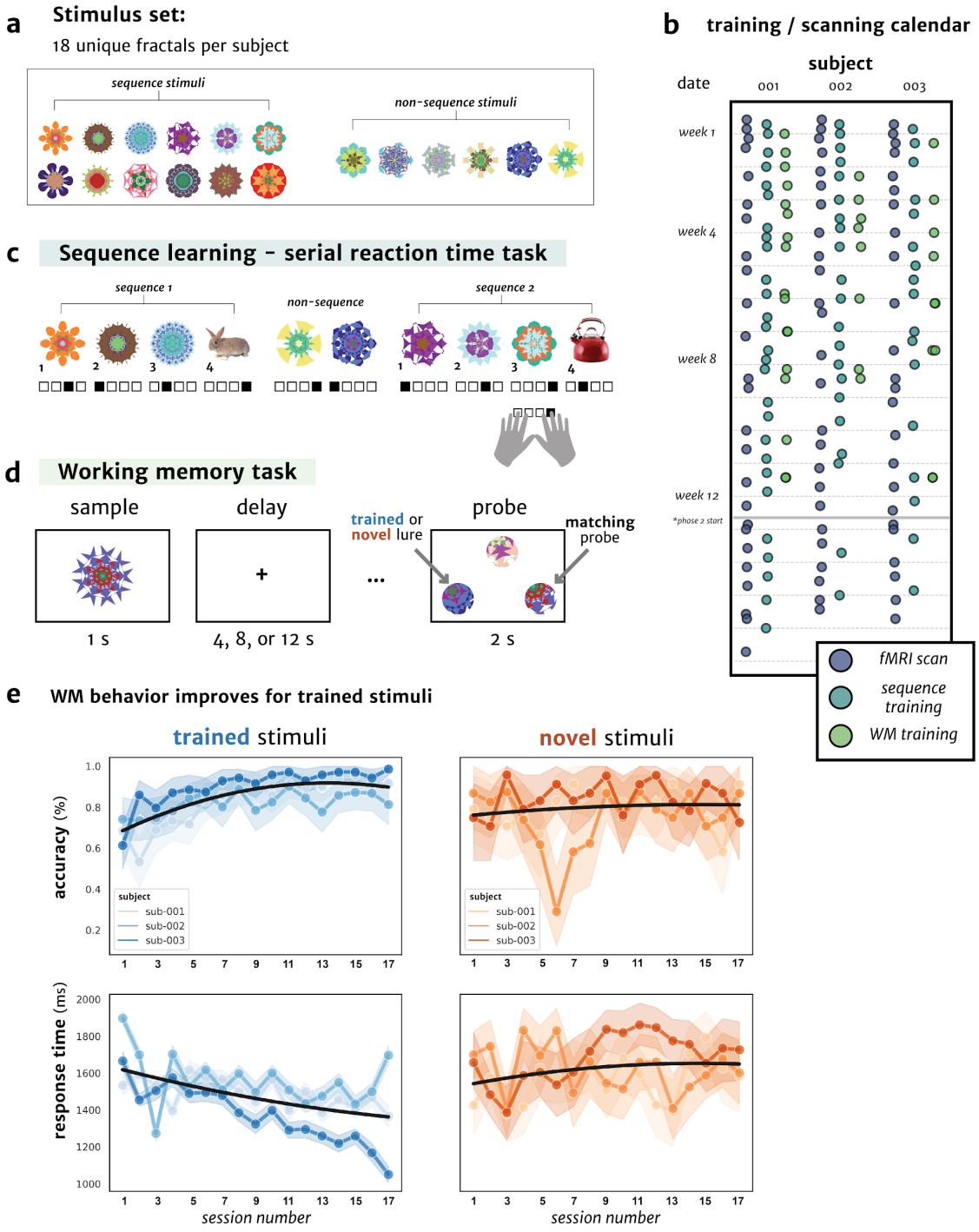


Figure 1. Longitudinal training across three months within individuals.
(a) Example set of 18 unique fractal stimuli assigned to a single participant for the in-scanner and at-home behavioral tasks. **(b)** Calendar of all of the MRI (purple) and at-home sessions (SRT - dark green, WM - light green) for each of the three participants over the four months of the study. During each MRI session, participants completed both the sequence learning and WM tasks. The at-home training sessions consisted of modified versions of each task (**Methods**). The present study analyzes the first 17 sessions, as afterwards new stimuli were added into the training set for each participant. **(c)** The serial reaction time (SRT) task, in which each of the 18 trained stimuli was associated

with one of four button responses. Of the 18 trained stimuli, 12 were part of 4 sequences that occurred with high probability (75%) in the SRT task, and participants learned the sequences over time (**SI Figure 1**). **(d)** The delayed three-alternative forced choice WM task, in which one fractal (trained or novel) was presented on each trial. After a jittered delay, participants indicated which occluded image matched the original sample. **(e)** WM task accuracy (top) and response time (bottom) improved across training (sessions 1-17) for trials with one of the 18 trained stimuli (*blue*), but not for trials with novel fractal stimuli (*orange*).

Divergent changes in mean WM delay activity within dorsal PFC

To determine if lPFC functioning changes across learning, we split the lPFC into six regions of interest (ROIs) along rostral-caudal (from the *frontal pole* to *precentral gyrus)* and dorsal-ventral (from the *superior frontal gyrus* to *inferior frontal gyrus*) axes (**Figure 2a**). This six-region parcellation was chosen to be homologous to a recent NHP study that recorded from multiple lPFC areas before and after WM training (Riley et al., 2018). We first tested for evidence of broad changes in WM delay activity over time. To test for changes in mean activity across entire ROIs, we considered two groups of voxels within each ROI. First, we examined whether peak activation in the WM delay period changes across sessions, which may reflect classical persistent activity during WM (Curtis & Sprague, 2021). To do this, we thresholded WM delay activity maps (collapsed across all delay lengths) for each participant and session at $t > 2.5$ and determined whether peak activation changes over training (**Figure 2b**, *top*). Second, we analyzed the mean activity of all voxels across each ROI, without any thresholding, to ask whether there are changes across an entire cortical region (including voxels with lower WM activity).

The magnitude of WM delay activity changed across training in two lPFC areas. First, the peak WM delay period activity in dorsal rostral PFC decreased across sessions (main effect of session number, mixed nonlinear model: $b = -0.031$, $t(45) = 2.71$, $p = 0.009$; **Figure 2b**, *top*), whereas the mean activity for all voxels in this area did not change over sessions ($b = 0.023$, $t(45) = 1.27$, $p = 0.21$; **Figure 2b**, *bottom*). Dorsal mid-lateral PFC showed the opposite pattern, with an increase in the mean activity across all voxels (main effect of session number across training, $b = 0.043$, $t(45) = 2.85$, $p = 0.006$; **Figure 2b**, *bottom*), but no change for peak activation ($b = -0.002$, $t(45) = 0.13$, $p = 0.89$; **Figure 2b**, *bottom*). No other ROIs showed training-related changes in either the peak WM delay activity or mean across all voxels (*p-values* > 0.1; **SI Figure 2**). However, this approach may obscure divergent changes that occur within specific populations of voxels with learning. We next directly tested whether individual voxels increased or decreased their activity over time using a voxelwise regression approach.

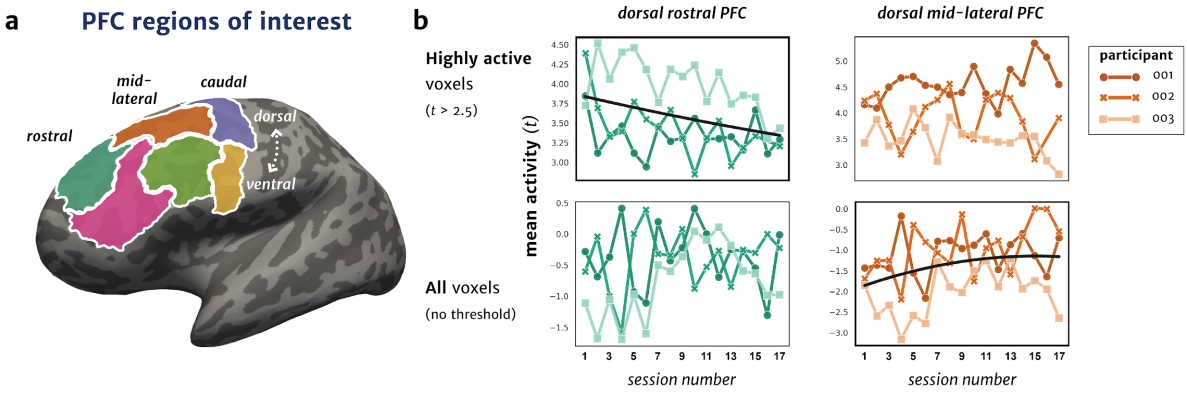


Figure 2. Mean WM delay activity changes in PFC across the course of learning.
(a) Six-region parcellation of the lateral PFC in an example participant's inflated left hemisphere. The lPFC was divided along a rostral-caudal and dorsal-ventral axis by combining smaller parcels from a multi-modal atlas of the cerebral cortex (Glasser et al., 2016). The parcellation was designed to be homologous to NHP electrophysiology studies (Riley et al., 2018), and guided by functional subdivisions of human lPFC (Badre & D'Esposito, 2009). **(b)** Top: Mean activity for each fMRI session during the WM delay period for reliably active voxels (within each session), thresholded at $t > 2.5$. The dorsal rostral PFC ROI (*green*), showed a mean decrease in WM delay activity across sessions. Bottom: Mean activity for all voxels in an ROI (unthresholded). The dorsal mid-lateral PFC ROI (*orange*) showed a mean increase in WM delay activity across sessions. For visualization, all ROIs with significant *b* parameters from nonlinear mixed models are indicated with a bolded plot border, along with the fitted nonlinear mixed model curve across sessions. No other ROIs showed a mean change in WM delay activity over the course of training.

More cortical territory in PFC is recruited for WM delay activity across learning

Populations of voxels involved in WM maintenance may change their activity over training, as the stimuli and task become increasingly well-learned. For example, WM processing could become more "efficient" by recruiting less cortical territory. Or, more cortical territory could be engaged in representing and processing newly learned stimuli and task dimensions. To test these different predictions, for each voxel, we assessed the relationship between WM delay activity and training time with nonlinear regressions (**Figure 3a**). We tested whether a meaningful proportion of individual voxels within each frontal ROI show systematic changes in activity over training compared to chance (permutation testing, see **Methods**). A schematic of this voxelwise approach is shown in **Figure 3a**, allowing us to test whether populations of voxels show divergent increases or decreases in WM delay activity in each ROI with training—information that would be lost when averaging across voxels.

For all lPFC ROIs, a distributed group of voxels increased in WM delay activity with training. That is, in every ROI, a significant percentage of voxels showed increased WM delay activity across the 17 sessions compared to chance (**Figure 3b**; *dorsal rostral:* $p < 0.001$*, dorsal mid-lateral:* $p < 0.001$*, dorsal caudal:* $p = 0.01$*, ventral rostral:* $p = 0.007$*, ventral mid-lateral:* $p = 0.03$*, ventral caudal:* $p = 0.002$*;* permutation tests). The dorsal mid-lateral and ventral caudal PFC showed the largest percentage of voxels with increasing WM delay activity over months of training (~25% of voxels). Specific to dorsal caudal PFC, one group of voxels within this ROI exhibited increased activity over time ($p = 0.01$), whereas another distinct group of voxels exhibited decreased activity ($p = 0.032$). These observed changes across all of lPFC were specific to the WM delay period, as the encoding (sample) period instead showed widespread decreases in activity with training across all ROIs (**SI Figure 3**).

In summary, repeated task and stimulus exposure was most commonly associated with increased WM delay period activity in a distributed group of voxels across lPFC, suggesting that these areas are more involved in WM maintenance over training. However, this increased activity may stem from the development of selectivity for individual stimuli over time, or a non-specific WM maintenance process that conveys no item-level information content. Therefore, we next tested whether frontal voxels with increases in WM delay activity show a corresponding differentiation in activity between individual trained stimuli.

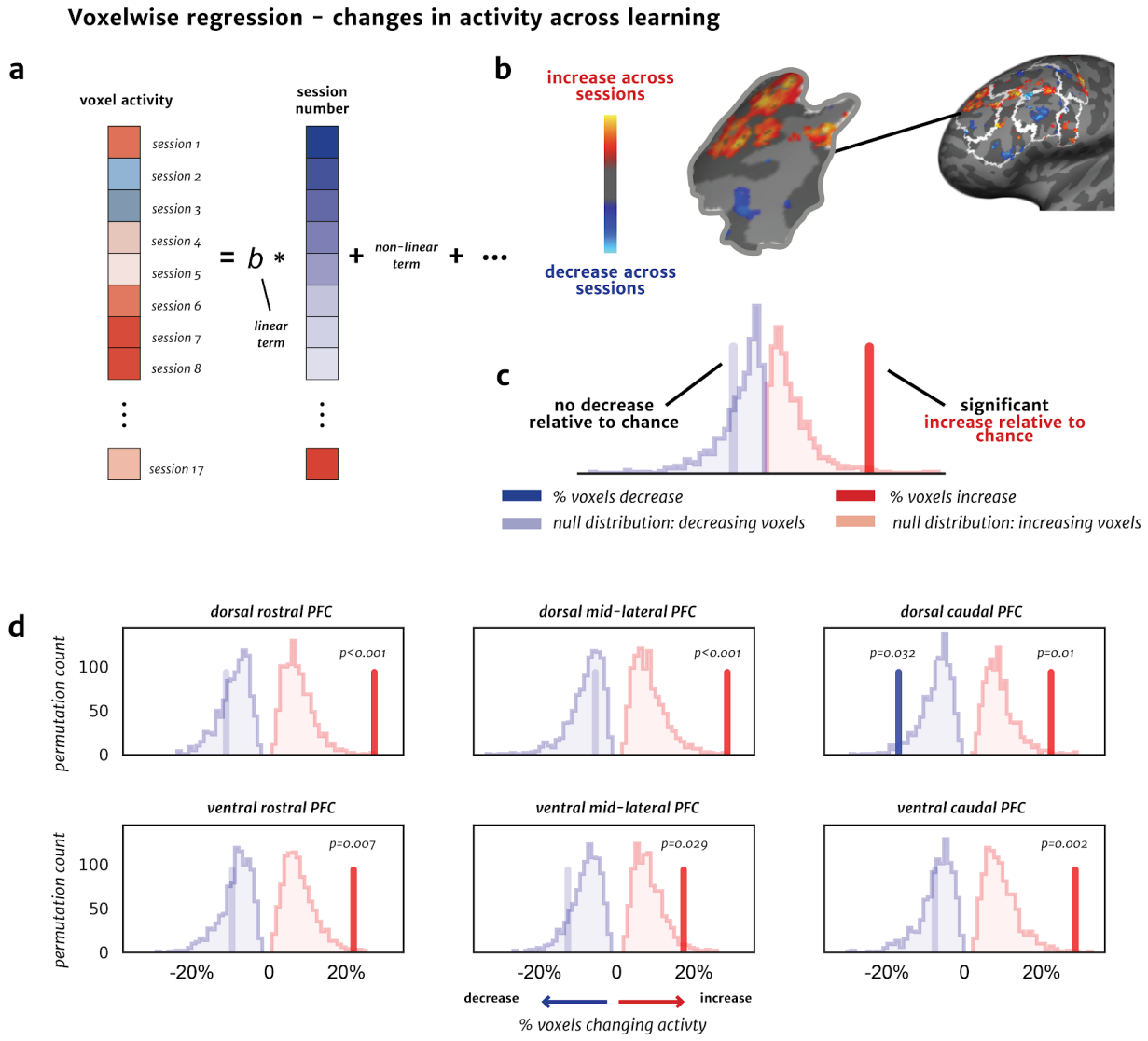


Figure 3. Distribution of WM delay activity patterns in PFC across the course of learning.
(a) Schematic of the linear term from the voxelwise regression approach, in which the mean WM delay activity from each voxel was regressed against a session number regressor in a nonlinear model (**Methods**). **(b)** Example *b*-parameter map (thresholded at $p < 0.05$) for one participant, the result of the linear term from the voxelwise nonlinear regression. Increases in activity across training are in red, with decreases in blue. **(c)** Percentage of voxels with increases (*red;* $b > 0$) or decreases (*blue;* $b < 0$) in activity across training (schematic). Significant changes over time are indicated by bolded vertical lines. Null distributions (created by permuting session number in the voxelwise regressions) are shown in light red and blue. **(d)** The percentage of voxels with significant changes in activity levels across training within each of the six lPFC ROIs. All ROIs show a significant proportion of voxels with an increase in activity, while only the dorsal caudal PFC also shows a significant proportion of activity decreases.

Changes in WM delay activity in ventral mid-lateral PFC correspond to increases in stimulus selectivity

What underlies the observed increase in WM delay activity with extensive training? We examined whether these changes (**Figure 3**) are associated with a corresponding increase in selectivity among the trained stimuli. We focused on the lPFC voxels that increased in WM delay activity across training for each participant, and we tested if these voxels displayed changes in stimulus selectivity. We generated a voxelwise selectivity index by analyzing single-trial WM delay activity profiles for each voxel across each participant's 18 trained stimuli. Then, two example voxels are highlighted in **Figure 4a** to show levels of WM delay activity for each of the trained stimuli early (*session 2)* versus late (*session 16*) in the 3-month training period. For each ROI, we tested whether the stimulus selectivity index (*F*-value) increased as a function of training when considering all voxels. Specifically, we used the voxelwise selectivity data in nested, nonlinear mixed models (**Figure 4b**). To test whether any selectivity changes are above and beyond what would be randomly expected over time, we created a distribution of null models (**Methods**). The nonlinear models show an increase in stimulus selectivity for the ventral mid-lateral PFC region ($p = 0.021$; **Figure 4b**, *right insets*), with no other ROIs showing a significant linear change over time (different *b*-parameter value from chance, P's > X).

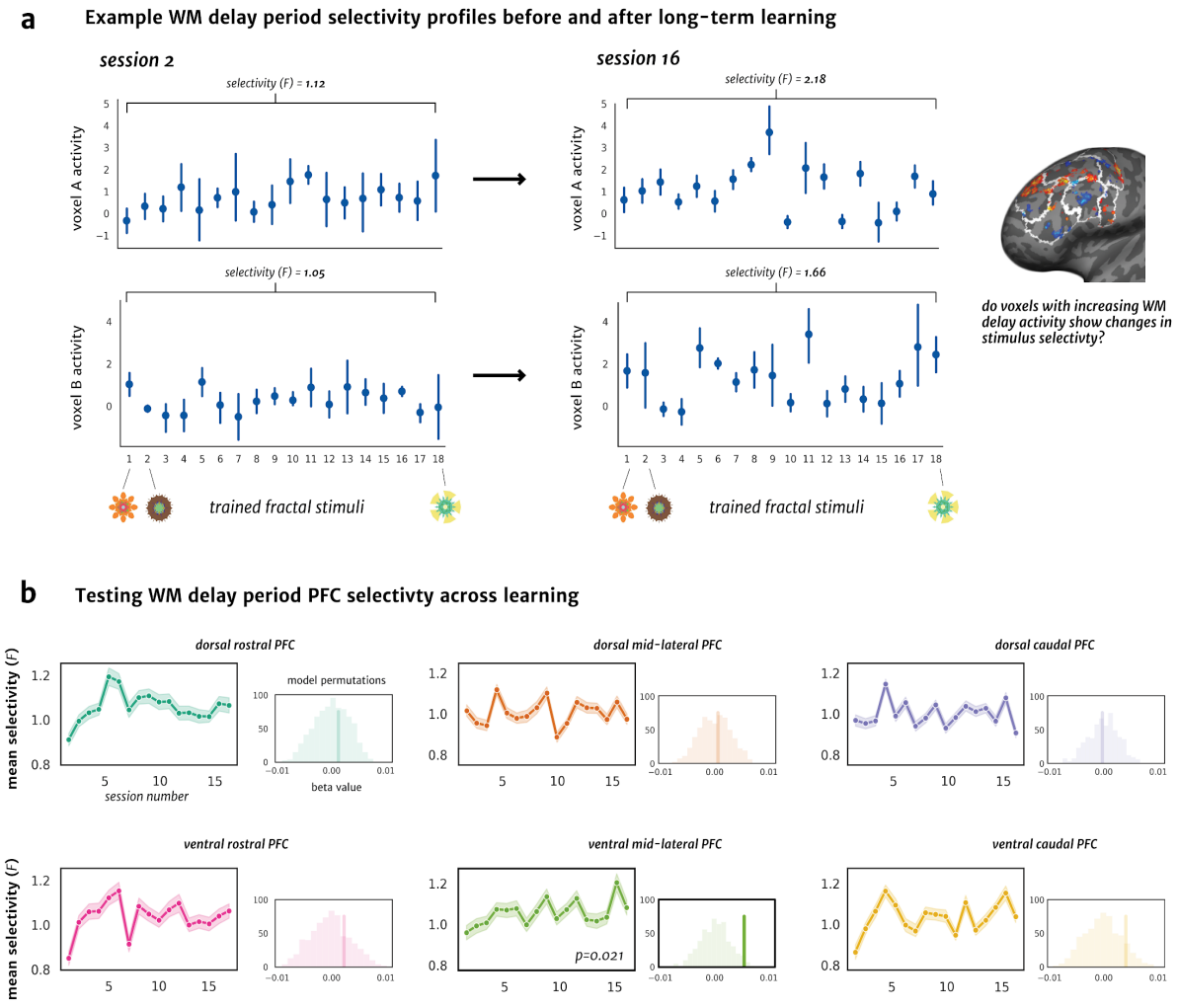


Figure 4. Increases in WM delay activity correspond to increases in stimulus selectivity across training among individual voxels in lPFC. (a) The mean WM delay activity of two example lPFC voxels for each of the 18 trained stimuli early (*left*, session 2) and late (*right*, session 16) into training, highlighting an increase in selectivity values (*F*-statistic) across the course of learning. **(b)** For each ROI, the left panel shows the mean selectivity index across sessions among all voxels with increasing WM delay activity. Shaded area represents a bootstrapped 68% CI. The right panel shows any significant selectivity increases across sessions (bold vertical line), as measured by the *b*-parameter of the nonlinear model. Null distributions of *b*-parameter values (histogram) from models with session number shuffled are shown in lighter colors. For visualization, all ROIs with significant *b* parameters compared to null distribution are indicated with a bolded plot border and bold vertical line.

Representational similarity patterns emerge for stimulus category, individual items, and sequence category in WM delay activity

After probing changes in activity at the single voxel level, we next tested whether the multivariate activity patterns across populations of voxels develop stimulus specificity over time. We took advantage of the extensive sampling in our dataset to test whether item-specific representational structures also appear in multi-voxel patterns of WM delay activity across the course of training. To do so, we employed a pattern similarity analysis framework (**Methods**). We estimated the similarity of representations across individual stimuli (matrices shown in **Figure 5; Methods**) by computing correlations between the WM delay period activity patterns for each stimulus. We then created several models to capture hypothesized levels of representational information (e.g., item or category level) and tested how well the observed similarity patterns matched the idealized models, producing a measure of representational "pattern strength" for each ROI in each session (**Methods**; **Figure 5a**). To determine if any pattern similarity effects were specific to the lPFC or would also be reflected in sensory areas, we examined patterns from early visual cortex (V1-V4) and the lateral occipital complex (LOC), a higher-order visual region.

First, we tested whether distinct representations of individual items in WM in lPFC or visual ROIs would emerge across training. We operationalized an *item-level* model for individual stimulus representations by testing for greater within-item pattern similarity (maintenance of the same trained stimulus across different trials, on-diagonal values in correlation matrix) compared to between-item similarity (maintenance of different trained stimuli, off-diagonal correlations), as schematized in **Figure 5b** (*left*). We focused on the six stimuli for each participant that were not part of regularly occurring sequences, in order to avoid capturing the possible restructuring of items in temporal sequences that may develop more integrated or differentiated representations over time (Sakai & Miyashita, 1991; Schapiro et al., 2012; Schlichting et al., 2015).

Pattern strength for the *item-level* model showed a significant increase over time in mid-lateral lPFC (*dorsal mid-lateral*: $b = 0.0004$, $t(46) = 2.34$, $p = 0.024$; *ventral mid-lateral*: $b = 0.0004$, $t(46) = 2.57$, $p = 0.014$; **Figure 5b**, *right*) and not in visual areas. That is, patterns of WM delay activity for individual trained items became more robust (reliable across trials) and differentiated from other trained stimuli across learning. To further test this effect, we also used a mixed linear model to compare the pattern strength for item-level selectivity in the first versus second half of sessions. When including all ROIs as levels of a categorical predictor in the model, mid-lateral PFC areas showed an interaction with learning time (*first* vs *second* half of sessions): *dorsal mid-lateral*: $\beta = 0.0056$, $t(45) = 2.66$, $p = 0.008$; *ventral mid-lateral*: $\beta = 0.0047$, $t(45) = 2.24$, $p = 0.026$. This analysis provides evidence for stronger item-specificity in patterns of lPFC delay activity across the course of training.

We next asked whether WM representations of all items show evidence of neural differentiation over time, or whether this is specific to trained stimuli. If the emergence of item-specific representations in lPFC is specific to trained stimuli, then activation patterns between trained stimuli should become less similar (as the items become more identifiable from each other) while those between novel stimuli should not reliably change. We operationalized this comparison with a *category-level* model which tested for an interaction with a decrease in pattern similarity between trained stimuli (that were not part of sequences) and no change similarity between novel

stimuli (off-diagonal correlations) as schematized in **Figure 5c** (*left*). There was a significant increase in pattern strength for the *category-level* model across sessions in dorsal caudal lPFC (*dorsal caudal*: $b = 0.0006$, $t(46) = 2.57$, $p = 0.013$; **Figure 5c**, *right*). This effect was driven by a decrease in the similarity between trained stimuli over time (*dorsal caudal*: $b = -0.001$, $t(46) = -3.62$, $p < 0.0017$) that was not observed between novel stimuli (*dorsal caudal*: $p = 0.74$). These pattern similarity analyses show a difference in the representations of trained and novel stimuli across learning, such that distinct representations of trained, but not novel, stimuli in WM emerge with learning.

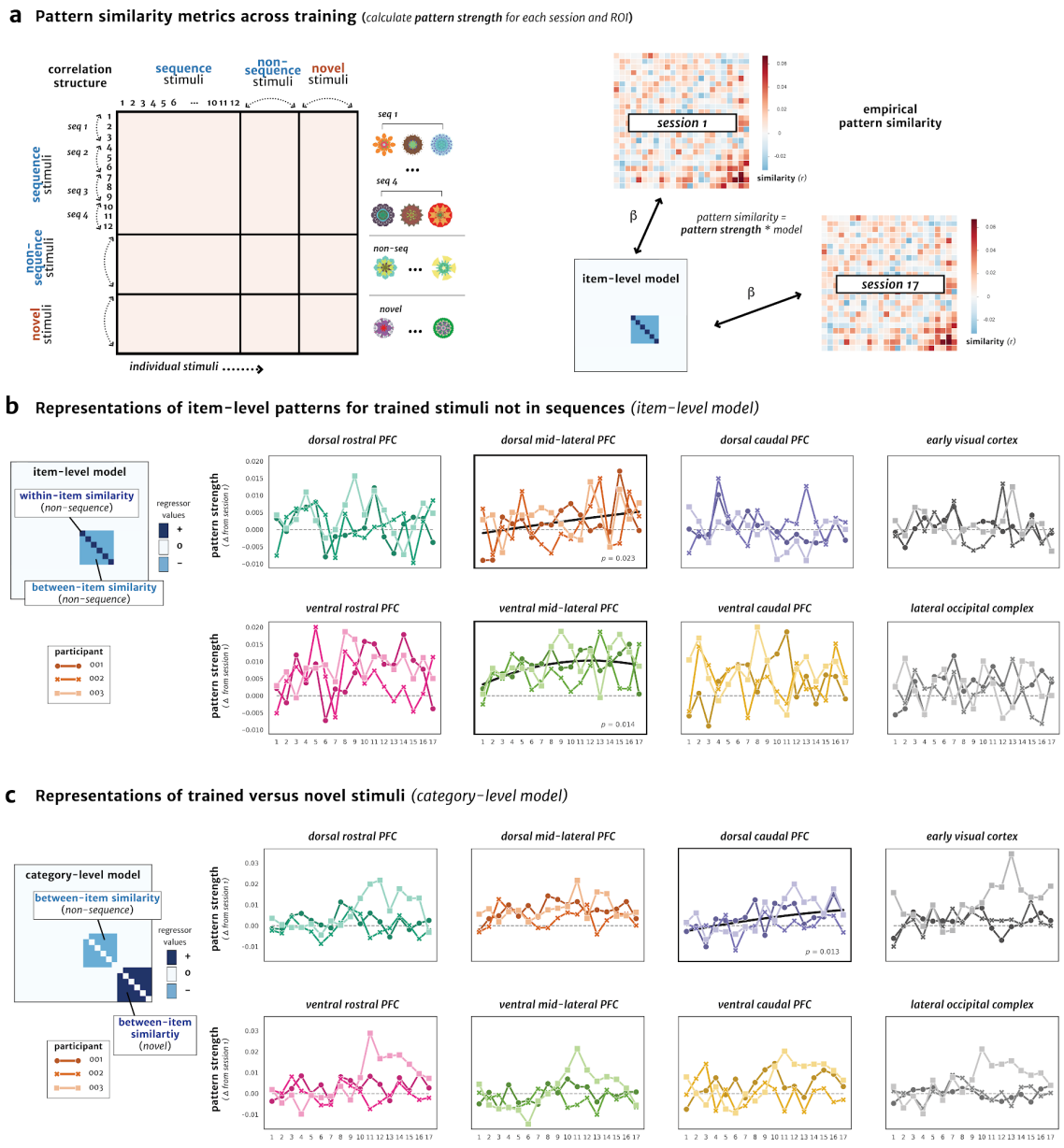


Figure 5. Emergence of representational similarity patterns for trained items in WM delay activity.
(a) Left: Schematic of a pattern similarity matrix of WM delay activity patterns across different stimuli. Right: Calculation of the pattern strength metric for each ROI and session by regressing a pattern model against the empirical pattern similarity data. **(b)** Left: Schematic of RSA framework for the *item-level* model with an interaction between on- (dark blue, positive values) versus off-diagonal (light blue, negative values) correlations among trained stimuli not in sequences as a measure of item-level representation. Right: Plots of the pattern strength across sessions for each ROI, as assessed by the model fit for the on- versus off-diagonal interaction on left. For visualization, all ROIs with significant changes in pattern strength across sessions are indicated with a *p*-value and bolded plot border, and pattern strength is plotted as a change from initial (session 1) baseline values. Each line represents one of the three individual participants. **(c)** Same as in **(b)**, but instead using the category-level model with a trained (light blue, negative values) versus novel (dark blue, positive values) stimulus off-diagonal interaction.

Finally, we tested whether associations learned in a distinct task context may influence WM maintenance processes even when they are not task-relevant. In parallel to the WM task, participants learned that a subset of trained stimuli formed high-probability temporal sequences in the SRT task (**SI Figure 1**). Based on classic studies of paired associate learning (Chen & Naya, 2020; Sakai & Miyashita, 1991) and multivariate representations that are altered by learning (Schapiro et al., 2012; Schlichting et al., 2015), we tested for shared representational structures across items in the same temporal sequence (higher similarity across items *within* the same sequence vs. *between* sequences, **SI Figure 4**) but found no effects for shared sequence-level patterns in any ROIs during WM maintenance (p-values > 0.05).

We then tested whether the organization of stimuli into temporal sequences in the SRT task may have resulted in a shared representation between stimuli that belonged to any sequence (regardless of sequence identity) which is distinct from items that were not part of a reliable temporal structure (non-sequence items). This coarse-level representation of sequence structure was operationalized with a *sequence category* model (**Figure 6a**, *left*). Pattern strength for this *sequence category* model showed a significant increase across sessions in caudal lPFC regions (*dorsal caudal*: $b = 0.0002$, $t(46) = 2.99$, $p = 0.004$; *ventral caudal*: $b = 0.0002$, $t(46) = 2.44$, $p = 0.019$; **Figure 6a**, *right*). This interaction was driven by a decrease in pattern similarity between sequence and non-sequence stimuli across sessions (*dorsal caudal*: $b = -0.0007$, $t(46) = -2.76$, $p = 0.008$; *ventral caudal*: $b = -0.0008$, $t(46) = -4.01$, $p = 0.002$). Across these analyses that consider associations in the SRT task, stimuli from learned sequences become more similar to each other over training, relative to stimuli not in sequences, specifically in caudal lPFC regions. These pattern similarity results suggest a categorical representation in which learned associations from LTM are reflected in WM delay activity, even when not directly tested in the WM task.

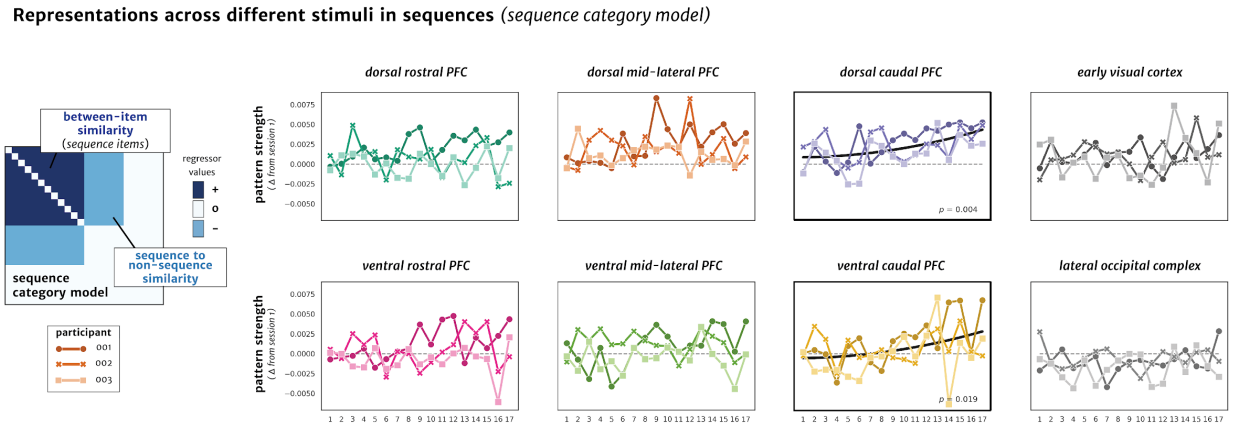


Figure 6. Emergence of pattern similarity for a categorical sequence representation in WM delay activity. Left: Schematic of the model matrix for the analysis of correlations for items within trained sequences (dark blue, positive values) compared to correlations of trained items not in sequences (light blue, negative values). Right: Plots of the pattern strength across sessions for each ROI, as assessed by the model fit for the *sequence category* model on the left. For visualization, all ROIs with significant changes in pattern strength across sessions are indicated with a *p*-value and bolded plot border, and pattern strength is plotted as a change from initial (session 1) baseline values. Each line represents one of the three individual participants.

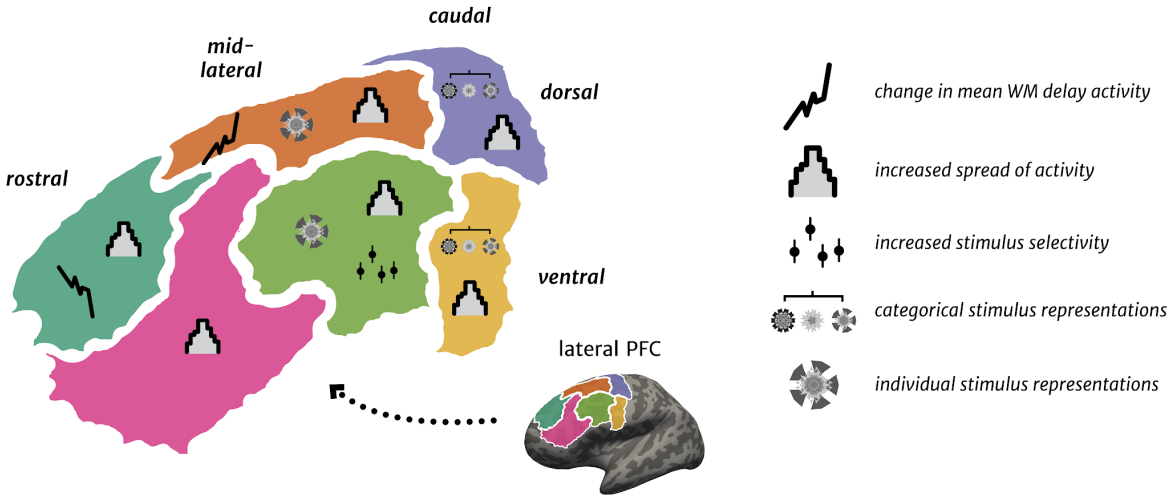


Figure 7. Summary of results.
Left: Each lPFC region, depicted with the observed training-related changes in the magnitude or pattern of WM delay activity. Right: Legend for the symbols depicting significant changes in WM delay activity levels, the voxelwise selectivity of the activity levels across stimuli, or multivariate representations (from pattern similarity analyses) for sequences and items in WM.

Discussion

Here, we aimed to determine how long-term learning influences lPFC neural representations for WM. Over three months, we extensively trained three human participants on a WM task and a sequence learning (SRT) task which both employed a unique set of complex, fractal stimuli. We found that the distribution and selectivity of lPFC WM delay activity changed across training: more cortical territory was recruited during the WM delay period with learning, and these activity changes coincided with increases in stimulus selectivity at the level of both individual voxels and multivariate patterns (**Figure 7**). Below we expand on the present results' implications for the role of lPFC during WM and interactions between LTM and WM.

lPFC - representations or processes?

Early NHP electrophysiological recordings from lPFC revealed neurons that respond to all phases of WM tasks: cue, delay, and response periods (Funahashi et al., 1990). Since then, neurons in NHP lPFC have been shown to encode both stimulus representations (Funahashi et al., 1989; Murray et al., 2017) and cognitive processes, including motor responses, rule learning, and executive control signals (Rigotti et al., 2013; Vallentin et al., 2012; Wallis & Miller, 2003). In contrast, human lPFC shows a relative absence of stimulus specific representations during WM (D'Esposito & Postle, 2015; Harrison & Tong, 2009; Leavitt et al., 2017; Serences, 2016) and human neuroimaging and lesion studies consistently point to lPFC mainly as a source of cognitive control signals (Chatham et al., 2014; Gazzaley & Nobre, 2012; Szczepanski & Knight, 2014). Thus, the role of lPFC function during WM has been unclear across studies. However, NHP and human studies are characterized by stark differences in training regimes before neural recordings take place (Berger et al., 2018; Birman & Gardner, 2016; Sarma et al., 2016). Therefore, we reasoned that differences in task and stimulus experience may underlie the discrepant conclusions about lPFC function. To directly test the influence of training on WM and lPFC function, we scanned participants across three months of repeated WM task and stimulus exposure. We provide novel evidence that extensive training facilitates stimulus specific representations in human lPFC during WM maintenance. The present results show long-term learning as a key influence over stimulus specific WM content in human lPFC, when this WM content is difficult to detect in human lPFC relative to visual areas (Bhandari et al., 2018; D'Esposito & Postle, 2015; Serences, 2016).

However, a small number of other notable studies also detect stimulus representations in human lPFC during WM, most likely by examining retinotopically organized areas in lPFC with visual orientation stimuli (Christophel et al., 2012; Ester et al., 2015) or using objects embedded in a high-level stimulus category task (Lee et al., 2013). Here, we specifically demonstrate that repeated stimulus and task exposure play a major role in facilitating r stimulus specific activity patterns in human lPFC, analogous to patterns more commonly found in NHP studies. There is also a potential dissociation in which initially highly active voxels may be engaged in more cognitive control processes for WM (Christophel et al., 2017; Serences, 2016; Sreenivasan et al., 2014), but a different group of voxels in the same area may develop stimulus-selectivity that only emerges with learning. Accordingly, these results point to training as a major factor in the debate over what information is encoded in lPFC activity during WM: general responses to task phases

(*processes*) are present without extensive training, while responses to individual stimuli (*representations*) are demonstrated in lPFC after long-term learning.

Implications for models of functional organization of lPFC

The lPFC is organized in a macroscale gradient along the rostral-caudal axis, both functionally (Badre & Nee, 2018; Koechlin et al., 2003) and anatomically (Goulas et al., 2014; J. A. Miller et al., 2021; Wagstyl et al., 2020). In this organization, more abstract representations are encoded more rostrally along the lPFC (Badre & D'Esposito, 2009), with middle frontal areas posited to sit "atop" such a hierarchical organization and provide top-down control signals during complex cognitive tasks (Badre & Nee, 2018; Duverne & Koechlin, 2017; Ito et al., 2017). Here, our data support a general rostral-caudal gradient in lPFC separating representational levels in WM: stimulus-specific representations were found in mid-lateral lPFC areas, categorical representations in the most caudal lPFC areas, and the rostral lPFC not showing such stimulus or category level representations (**Figure 7**). This separation of WM representations along a rostral-caudal lPFC axis was only present after repeated stimulus and task exposure, suggesting that learning may use this existing organization in order to easily separate similar but related representations for successful WM behavior. While here we show WM stimulus patterns emerge in micro-anatomically similar areas to representations found in NHP recordings (e.g., Brodmann's areas 9/46d, 9/46v; (Petrides, 2005)), functional homology is often less similar across NHP and human lPFC. For example, lesions of caudal precentral areas in human lPFC cause deficits in spatial WM that mirror damage to more anterior, mid-dorsal lPFC areas in NHPs (Mackey et al., 2016). Where WM representations are actively maintained may then shift in location based on learning and task demands. Here, we show that such learning drives mid-lateral lPFC regions most often described as a "controller" of task activity to maintain representations of a given WM stimulus. Future work using longitudinal paradigms in NHP studies might also clarify the importance of training, spatial scale, and species differences on WM maintenance processes (Badre et al., 2015; Milham et al., 2018; Song et al., 2021).

Plasticity of the PFC

The lPFC is critical for flexible cognition and multiple theories consider the lPFC to have high plasticity, with activity patterns and representations changing based on task demands over time (Duncan, 2001; Woolgar et al., 2011). However, these patterns of adaptation from task learning have not been systematically tracked over time in human lPFC. Some human neuroimaging studies have employed forms of WM training as a route to improve WM and cognition more broadly. These early studies found activation increases in frontal and parietal cortex after WM training (Klingberg, 2010; Olesen et al., 2004), but recent aggregations of WM training studies detail a broader mix of activation changes, with decreases for studies with shorter training times (~minutes-hours) and increases for longer training (~days-weeks) (Buschkuehl et al., 2012, 2014). However, consistent conclusions are elusive as these studies do not collect neuroimaging data at wide a variety of time points to track learning across time, nor do they examine the effects of stimulus experience and context on the neural mechanisms of WM maintenance.

Recently, NHP electrophysiology studies have observed changes in the selectivity and magnitude of both single-unit and population spiking during WM (Dang et al., 2021; Meyer et al., 2011; Qi et al., 2019). These effects were greatest in mid- and anterior dorsal areas of lPFC, resembling the emergence of stimulus-selective activity patterns that we observe here in mid-lateral lPFC. The long-term plasticity in mid-lateral PFC that we observe here is likely enabled by several factors that give the region a high propensity for flexible representations: long-range anatomical connections (Chaudhuri et al., 2015; Y. Wang et al., 2021), status as a hub between cortical networks (Bertolero et al., 2018; Fornito et al., 2019), and a late anatomical development (Garcia-Cabezas et al., 2019; Garcia et al., 2018).

Influence of LTM on WM

In classic theories, WM and LTM are thought to rely on both different brain areas and neuronal mechanisms for memory storage (Squire & Zola-Morgan, 1991; Warrington & Shallice, 1969; Wickelgren, 1969). Thus, the neural substrates for WM are most often studied without the consideration of any longer term learning and memory effects. When the influence of training is considered for WM performance, better WM is observed for familiar, complex stimuli such as Pokémon (Xie & Zhang, 2017), meaningful human faces (Asp et al., 2021; Jackson & Raymond, 2008), and trained geometric shapes (Blalock, 2015). Our findings suggest that these experience-dependent WM effects are underpinned by malleability of the cortical representations supporting WM across the course of learning. Specifically, our results show that selectivity for individual stimuli increased in human lPFC activity patterns across months of training. We also found shared representations across stimuli that were part of temporal sequences, consistent with the emergence of a "categorical" representation separating items based on their temporal properties in the SRT task. The categorical sequence representation may have emerged as a function of memory consolidation and repeated practice, leading to a semantic code for stimuli occupying the same class of patterns over time (sequence stimuli) versus a distinct class of non-sequence stimuli(Binder & Desai, 2011; Eichenbaum, 2017; Nadel & Moscovitch, 1997; Sommer, 2017; Winocur & Moscovitch, 2011). Individual stimulus and categorical representations also emerged in different areas of lPFC, suggesting long-term learning changed activity patterns along relevant functional axes of lPFC organization (**Figure 7**). Altogether, the results provide key evidence not only that LTM can share representational formats with WM (Beukers et al., 2021; Lewis-Peacock & Norman, 2014; Nee & Jonides, 2011; Oberauer, 2009), but that long-term learning *changes* how information is represented in WM, even when learned associations are not behaviorally relevant for WM. .

Considerations for future studies of WM

Here we demonstrate effects of long-term learning on the cortical substrates for WM maintenance in human lPFC that prompt a larger reconsideration of how WM studies are conducted and interpreted. That is, how neural circuitry supports WM is shaped by prior experiences and learning, which can lead to drastically different conclusions on the role(s) of lPFC in WM, depending on when brain recordings take place relative to prior experience. This timeline of learning is especially important to consider because neuronal ensembles in lPFC demonstrate a remarkable flexibility in activity magnitude, timing, and dimensionality based on task demands (Dang et al., 2021; E. K. Miller & Cohen, 2001; E. K. Miller & Fusi, 2013; Stokes et al., 2013;

Wasmuht et al., 2018). Altogether, these data show how long-term learning sculpts neural representations during working memory, suggesting that an array of cognitive processes and their associated neural circuitry are likely to be transformed by prior experience.

Methods

Data and Code Availability

All neuroimaging data will be openly available in the Brain Imaging Data Structure format ((Gorgolewski et al., 2016); https://bids.neuroimaging.io/) on the OpenNeuro platform upon publication (openneuro.org). Analysis and processing code to reproduce the present results, along with the stimuli, presentation code, and behavioral data may be found on Open Science Framework (OSF) : osf.org

Human participants

The three study participants were all healthy, adult volunteers. Because of the large amount of MRI data collected and intensive nature of the behavioral training involved, all participants were members of the research team who completed the study over the same time period. One participant was a 34-year-old female (sub-001), one was a 25-year-old male (sub-002), and one was a 37-year-old female (sub-003). The University of California, Berkeley Committee for the Protection of Human Subjects (CPHS) approved the study protocol and no participants reported any contraindications for MRI.

Study design and stimuli

The study was designed to investigate WM behavior and neural representations across a large amount of training on a specific set of stimuli and tasks. To accomplish this, we assigned each participant a unique set of 18 fractal images as their set of trained stimuli. Each image was an algorithmically-generated fractal consisting of multiple colors, and the 18 images for each participant were balanced according to the primary color group of each image (determined using a k-means clustering algorithm on each fractal image in the *sklearn* Python package: https://scikit-learn.org/). These fractals were chosen because they are visually complex, approximately uniform in size, cannot be easily verbalized, have no pre-existing meaning, and similar stimuli have been used in NHP electrophysiology studies of the neural basis of learning (Ghazizadeh et al., 2018; Kim et al., 2015; Sakai & Miyashita, 1991). Because the study participants were also on the research team, we avoided participants gaining any foreknowledge of their training set by generating thousands of initial images and randomly selecting each training set from among these images. Thus, each participants' first exposure to their training set occurred during the first scanning session. The unique 18 stimuli for each participant were then used for all of the following fMRI and behavioral training sessions, with additional novel stimuli randomly selected each session from the broader set of fractals. Of the 18 fractal stimuli in each participant's training set, 12 were randomly assigned to be part of four sequences in the SRT task, with each sequence consisting of three fractals and an object image. The sequences were not explicitly instructed and were learned over time as part of a serial reaction time (SRT) task (though all participants had knowledge of the sequence manipulation). All tasks were programmed using *Psychtoolbox* functions (Brainard, 1997; http://psychtoolbox.org/) in Matlab (https://www.mathworks.com/), and stimuli were presented on a plain white background [RGB = 255,255,255].

Longitudinal training

Across the course of 15 weeks, each participant underwent 24-25 total sessions of fMRI scanning. In the present work, we analyze the first 17 of these fMRI sessions (*Phase 1*) for each participant which took place over ~3 months (13 weeks) of training. In a second study phase (*Phase 2*)of ~3 additional weeks, more fractal stimuli were added into the training set (**Figure 1c**), but the results from this phase of the experiment are not reported here. Over the first week, four scans were conducted to ensure that the initial exposure to the tasks and stimuli would be highly sampled. fMRI scanning during subsequent weeks occurred at a rate of approximately 1-2x per week (depending on participant and scanner availability).

To facilitate learning, at-home behavioral training was implemented multiple times per week across the course of the study (**Figure 1c**), where Participants completed versions of the WM and sequence learning tasks on home laptop testing setups. Most sessions were completed at the same location for each subject, with a small number completed elsewhere (when traveling, for example). The at-home WM task training data can be found on Open Science Framework ().

Working memory task

Participants completed a three-alternative forced choice delayed recognition task in each scanning and at-home WM training session (**Figure 1a**). Stimuli included the 18 fractals from the participant's training set, along with 6 novel fractal images, which were randomly selected each session. On each trial, a single WM sample stimulus (600 x 600 pixels) was presented in the center of a screen for a 0.5 s encoding period. A fixation cross was then presented for a jittered delay period of 4, 8, or 12 s, with the goal of facilitating WM maintenance processes. A probe display then appeared for a response window of 2 s. The probe display comprised three occluded sections of fractal images (⅙ area of each image) at an equal distance from the center of the screen. Each probe image was masked within a gaussian window of FWHM at ~⅙ the image size. Participants responded via one of three button presses to indicate which probe image segment matched the stimulus from the beginning of the trial. A fourth button could be used to indicate a response of "I don't know.". A sample-matching fractal image was always present in the probe display. One of the other probe stimuli was always a novel (untrained) fractal image randomly selected from the same color group as the sample fractal image. The third probe image was either a novel fractal (50% of trials) or a lure from the set of trained fractal images (50% of trials). The masked section of the fractal images was in the same location for each probe image and randomly chosen from nine different areas on each trial, and the probe position was counterbalanced across trials within a block (**Figure 1a**). After each trial, there was a jittered intertrial interval (ITI) sampled from an exponential distribution (mean = 4 s, range = 1 - 9 s).

In the scanning sessions, participants completed four blocks of 24 trials, with each trained and novel fractal image presented as the WM sample stimulus once per block, in random order. Each delay length occurred in random order and equally often within a block. For the at-home WM training sessions, participants completed two blocks of 24 trials (**Figure 1c**). The in-scanner display was a back-projected 24 in. screen (1024 x 768) for an approximate ~47 cm viewing distance, while for at-home training sessions participants used laptop screens of sizes 13.3 in. (1440 x 900) [sub-001], 13.3 in. (2560 x 1600) [sub-002], and 12.5 in. (1920 x 1080) [sub-003].

Serial reaction time task

In addition to the WM task, participants completed a serial reaction time (SRT) task before the WM task in each scanning session and during at-home training sessions. This task served to repeatedly expose participants to statistical regularities amongst the trained stimuli, in the form of temporal stimulus sequences. During this task, participants made button presses in response to each stimulus. The stimulus set consisted of the same 18 fractal stimuli shown in the WM task as well as six objects (three animals and three tools) for a total of 24 stimuli. The SRT task consisted of two phases: an initial phase in which stimulus-response mappings were learned), followed by a second phase during which stimulus sequences were present.

The first section of SRT task was implemented in the first two sessions of the study (one fMRI session followed by an at-home behavioral session) during which participants were trained to criterion to associate each of the stimuli with one of four button press responses. Participants were first exposed to their stimulus set during their first scanning session. During every block, each of the 24 stimuli were shown once in a randomized order, with no explicit sequence information present (during the first two sessions). Each stimulus was presented on the screen for 2.3 seconds (followed by a blank screen of .7 s between stimuli) with four response options shown as black squares below the stimulus (corresponding to the middle finger of the left hand, ring finger of the left hand, ring finger of the right hand, and middle finger of the right hand). During the first two blocks of the first scanning session, the correct response was highlighted (square corresponding to the response was shown in red instead of black) to allow participants to view the correct response and facilitate learning. Thereafter, participants completed 10 more blocks during which the correct response was not shown but feedback was provided (when a correct response was made the square turned blue and incorrect responses were indicated by the selected option turning red with feedback lasting for 200 ms). After the first scanning session, participants performed an at-home session to ensure the learning of stimulus-response mappings. Participants completed a minimum of five blocks of the task, and continued until a criterion of 80% accuracy at the item-level was reached (>=80% of correct first responses for all stimuli across all blocks; 7 - 15 blocks of training were required to reach criterion). The stimulus-response mappings remained constant throughout the study.

After the completion of training to criterion, temporal sequences of stimuli were embedded in the SRT task, beginning in the second fMRI session. Of the 24 trained stimuli (18 fractals and six objects), 16 stimuli were assigned to form four distinct sequences, with each sequence containing three fractals followed by an object (**Figure 1b**). As in the initial section of this task, each stimulus was shown once during each block (set of 24 trials) and the four response options were indicated below the stimulus as four black squares. Participants were instructed to press the appropriate button for each stimulus. Each stimulus was shown for 1.95 s (fMRI sessions) or 1.8 s (behavioral sessions) followed by a blank screen for 400 ms. Sequences were presented in a probabilistic manner, such that three of the four sequences were presented in an intact fashion in each block and each sequence was intact on 75% of blocks in each session (i.e. in 12/16 blocks during fMRI sessions). In each block, the order of the presentation of stimuli was randomized with the exception of the presentation of the three intact sequences. Stimuli from the non-intact sequence (one sequence per block) were presented in a random order with the stipulation that at

least two stimuli separated the non-intact sequence stimuli. Feedback was provided throughout the experiment as described above in the training to criterion phase. The fMRI sessions contained 18 blocks of the SRT task and the at-home behavioral sessions consisted of 26 blocks. Stimuli were presented in a randomized order (no sequence information was present) during the first two blocks of each session which served to acclimate participants to the task.

Object-selective functional localizer task

Functional localizer scans were collected during two separate fMRI sessions for each participant, which occurred after sessions 1 and 5 for sub-001, sessions 1 and 15 for sub-002, and sessions 5 and 14 for sub-003. Participants performed a one-back task while viewing blocks of animals, tools, objects, faces, scenes, and scrambled images. All images were presented on phase scrambled backgrounds. Each block lasted for 16 s and contained 20 stimuli per block (300 ms stimulus presentation followed by a blank 500 ms inter-stimulus interval). Two stimuli were repeated in each block and participants were instructed to respond to stimulus repetitions via button press. Each scan (three scans per session) contained four blocks of each stimulus class, which were interleaved with five blocks of passive fixation.

fMRI acquisition

All neuroimaging data were collected on a 3 Tesla Siemens MRI scanner at the UC Berkeley Henry H. Wheeler Jr. Brain Imaging Center (BIC). Whole-brain Blood Oxygen Level-Dependent (BOLD) fMRI (T_2^*-weighted) scans were acquired with a 32-channel RF head coil using a 2x accelerated multiband echo-planar imaging (EPI) sequence [repetition time (TR) = 2 s, echo time = 30.2 ms, flip angle (FA) = 80°, 2.5 mm isotropic voxels, 52 slices, matrix size = 84 x 84]. Anatomical MRI scans were collected at two timepoints across the study and registered and averaged together before further preprocessing. Each T_1-weighted anatomical MRI was collected with a 32-channel head coil using an MPRAGE gradient-echo sequence [repetition time (TR) = 2.3 s, echo time = 3 ms, 1 mm isotropic voxels]. For each scan, participants wore custom-fitted headcases (caseforge.com) to facilitate a consistent imaging slice prescription across sessions and to minimize head motion during data acquisition.

In each 2-hr scanning session, participants completed the following BOLD fMRI scans: (1) 9 min eyes-closed rest run, (2) three 9 min runs of a 1-back stimulus localizer, (3) three 6 min runs of the SRT task, (4) 9 min eyes-closed rest block, (5) 9-min stimulus localizer block, (6) four 6 min runs of the WM task. The present work focuses on the WM task. In the stimulus localizer scans, participants completed a 1-back task with a slow, event-related design optimized for obtaining single-trial multivariate representations (Zeithamova et al., 2017) (results not reported here).

fMRI preprocessing

Preprocessing of the neuroimaging data was performed using fMRIPrep version 1.4.0 (Esteban et al., 2018), a Nipype (Gorgolewski et al., 2017) based tool. Each T1w (T1-weighted) volume was corrected for INU (intensity non-uniformity) using *N4BiasFieldCorrection* v2.1.0 (Tustison et al., 2010) and skull-stripped using *antsBrainExtraction.sh* v2.1.0 (using the OASIS template). Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (Zhang et al., 2001) (FSL v5.0.9).

Functional data was slice time corrected using 3dTshift from AFNI v16.2.07 (Cox, 1996) and motion corrected using mcflirt (Jenkinson et al., 2002) (FSL v5.0.9). This was followed by co-registration to the corresponding T1w using boundary-based registration (Greve & Fischl, 2009) with 9 degrees of freedom, using flirt (FSL). Motion correcting transformations and BOLD-to-T1w transformation were concatenated and applied in a single step using *antsApplyTransforms* (ANTs v2.1.0) using Lanczos interpolation. Many internal operations of FMRIPREP use Nilearn (Abraham et al., 2014), principally within the BOLD-processing workflow. For more details of the pipeline see https://fmriprep.readthedocs.io/en/latest/workflows.html. Finally, spatial smoothing was only performed in a 4mm FWHM kernel along the cortical surface (https://github.com/mwaskom/lyman/tree/v2.0.0) for the mean univariate activity analysis (**Figure 2**), while all other analyses used unsmoothed data.

Region-of-Interest (ROI) selection

To generate cortical surface reconstructions, the T_1-weighted anatomical MRIs were processed through the FreeSurfer (https://surfer.nmr.mgh.harvard.edu/) *recon-all* pipeline for gray and white matter segmentation (Dale et al., 1999; Fischl, Sereno, & Dale, 1999). To construct the lPFC ROIs, we sampled a recent multimodal areal parcellation of the human cerebral cortex (Glasser et al., 2016) onto each participant's native anatomical surface via cortex-based alignment (Fischl, Sereno, Tootell, et al., 1999). We combined these smaller parcels on the surface into six different lPFC ROIs, with two splits along the rostral-caudal axis and one split along the dorsal-ventral axis (**Figure 2b**). The caudal lPFC ROIs fall along the precentral sulcus and gyrus, with the most rostral ROIs ending in frontopolar cortex around the anterior ends of the inferior and superior frontal sulci. The split between dorsal-ventral ROIs roughly falls along the posterior middle frontal sulci, analogous microstructurally to the principal sulcus of macaques (J. A. Miller et al., 2021; Petrides, 2019), and the ROIs are bounded dorsally by the superior frontal gyrus and ventrally by the inferior frontal gyrus. This lPFC division into six areas was designed to align with NHP electrophysiology studies recording from multiple frontal cortex regions (Riley et al., 2018).

We also constructed two visual ROIs in order to determine if effects were specific to lPFC or also generalized to lower and higher-order visual areas. An early visual cortex ROI combined visual cortical areas V1-V4 for each participant, defined from aligning a probabilistic visual region atlas (L. Wang et al., 2015) onto each subject's native cortical surface using cortex-based alignment (**Figure 5a**). A higher-order visual ROI for the lateral occipital complex (LOC) was defined from a separate category localizer scanning session [block-level general linear model (GLM) with a contrast of responses of objects > scrambled objects]. Voxel responses were thresholded at $p < .0001$ and the ROI was restricted to voxels reaching this statistical threshold on

the lateral surface of the occipital cortex and the posterior portion of the fusiform gyrus (Schwarzlose et al., 2008) .

Mean WM delay activity across training

We constructed a separate event-related GLM in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/) for each participant and session in order to compare activity levels for each voxel across training. Separate boxcar regressors were constructed for the encoding (0.5 s), delay (4, 8, or 12 s), and probe (2 s) periods of the WM task, and all regressors were convolved with a standard double-gamma hemodynamic response function (HRF). Separate task event regressors were created for trained and novel fractals. For the session-level GLMs, all four WM task runs in each session were concatenated with the *spm_fmri_concatenate* function. Six rigid-body motion parameters were included as nuisance regressors, along with high-pass filtering (HPF) of 128s to capture low-frequency trends as implemented in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/). Voxelwise *t*-statistic maps were then calculated for WM delay (delay > fixation) periods, selecting regressors for trials across all three delay lengths. We analyzed changes in mean WM delay activity over learning with nonlinear mixed models using mean activity in each ROI as the outcome variable and session number as the predictor (*Statistical methods*). These analyses were performed in two broad groups of voxels: (1) for the mean activity of voxels within the peak activation for each ROI (thresholding the maps for each participant and session at $t > 2.5$) and (2) for the mean activity of all voxels across each ROI without any thresholding.

Voxelwise regression analysis (recruitment of voxels across training)

To ask whether voxels showed changes in activity across training, we performed voxelwise nonlinear regressions on the *t*-statistic values from the above GLMs (*Mean WM delay activity across training*) across sessions (**Figure 3a-c**). Separate voxelwise models were run on WM encoding and delay period activation to characterize changes in each phase of the WM task separately. For each participant and lPFC region, this regression generated a voxelwise *b*-statistic (linear term of quadratic model, see *Statistical methods*), with positive values indicating an increase in activity across sessions and negative values a decrease in activity across sessions. After thresholding the voxelwise *b*-statistic maps at $p < 0.05$, we then calculated the proportion of voxels in each ROI showing an increase or decrease in activity across sessions and averaged this value across participants. This generated a measure of how many voxels in an ROI change their activity over time, without requiring precise overlap of the specific voxels showing changes across participants. To determine if the proportion of voxels showing an increase or decrease in activity across sessions was different than chance ($p < 0.05 / 2 = 2.5\%$ false-alarm rate for increases or decreases), we constructed permuted null distributions of the proportion of increasing and decreasing voxels in each ROI. In each of 1,000 permutations, session number was randomly shuffled, the regression onto activity across sessions was re-computed, and the proportion of voxels showing increases and decreases in activity (mean across participants) was stored to create null distributions. *P*-values were then derived by comparing the actual proportion of increasing and decreasing voxels across participants (dark lines in **Figure 3d**) to the permuted null distributions.

Stimulus selectivity metric and analyses

In order to determine if WM delay activity showed preferences for any specific fractal stimuli we obtained single-trial level voxelwise activity maps by constructing separate least-squares-all (LSA) GLM for each run, session, and participant (Mumford et al., 2012). Here, GLMs were constructed separately for each run in order to estimate pattern similarity between different runs, so that correlation measures aren't confounded by temporal autocorrelation within each functional scan (Mumford et al., 2014; Zeithamova et al., 2017). In each run-level GLM, the WM delay period events for each of the 24 unique stimuli were modeled as separate boxcar regressors (collapsed across delay lengths) and convolved with a HRF. The combined WM encoding (0.5 s) and probe (2 s) events were included as nuisance regressors, again split by trained and novel stimuli. Six rigid-body motion parameters were also included as nuisance regressors, along with high-pass filtering (HPF) of 128s to capture low-frequency trends.Voxelwise *beta*-statistic maps from each trial were then used in the selectivity and pattern similarity analyses.

To determine if changes in lPFC activity show selectivity for the trained stimuli across training, we calculated a voxelwise selectivity index (among voxels that increased in WM activity across training) of WM delay activity for every session, lPFC region, and participant. Analogous to stimulus selectivity measures from electrophysiology studies (Naya et al., 2001; Wirth et al., 2003), an *F*-statistic was calculated for each voxel using WM delay activity levels (*beta* estimates) across the 18 unique trained stimuli in a repeated-measures ANOVA (with each of the four runs in the WM task as the repeated measure, **Figure 4a**). To determine if lPFC regions showed changes in selectivity across training, we implemented nested mixed nonlinear models (see *Statistical methods*) with selectivity as the outcome variable and session number as the predictor. Separate models were constructed for each ROI and data from every voxel was included as a nested variable within the participant (subject-level) variable. We used every voxel from the ROI to be more sensitive to detect changes across training than by using the mean alone, noting that the degrees-of-freedom were inflated because of correlations between voxels. Accordingly, we assessed the significance of an effect of session number (training) on selectivity using permutation testing. A null distribution of the relationship between session number and selectivity was created by shuffling the session number regressor in each of 1,000 permutations and re-computing the relationship between selectivity and session number. The *b*-statistic from the actual model was then compared to the null distribution of *b*-values for each ROI (**Figure 4b**).

Representational similarity analyses

To obtain measures of pattern similarity of the fMRI responses in each ROI across conditions, we applied a multivariate noise decomposition algorithm to the single-trial WM delay period responses (Walther et al., 2016). This process used the time-series of residuals from the LSA GLM for each run to account for noise variance within each ROI. Then, for each session, we calculated cross-validated (between-run) correlations between the trials for all stimuli (18 trained, 6 novel fractals). Correlation values were Fisher-z transformed, and then the mean of the between-run correlations generated a representational similarity or correlation matrix (**Figure 5a**). One total run across all sessions and participants was removed from calculation of between-run correlations because of a visual MR artifact. To test for distinct representational structures in WM delay period patterns, we operationalized each of four potential representations as specific predictors of pattern

similarity and then analyzed how the strength of each model changed across training. Each representational structure was coded using values of (*1, -1*) for specific stimulus pairs, with negative values weighted such that the regressor values summed to zero. After constructing, these values were then used as predictors of the similarity values (Fisher z-transformed pearson correlation), resulting in a model fit ("pattern strength") for each representational structure. This procedure was performed for each session, participant, and ROI.

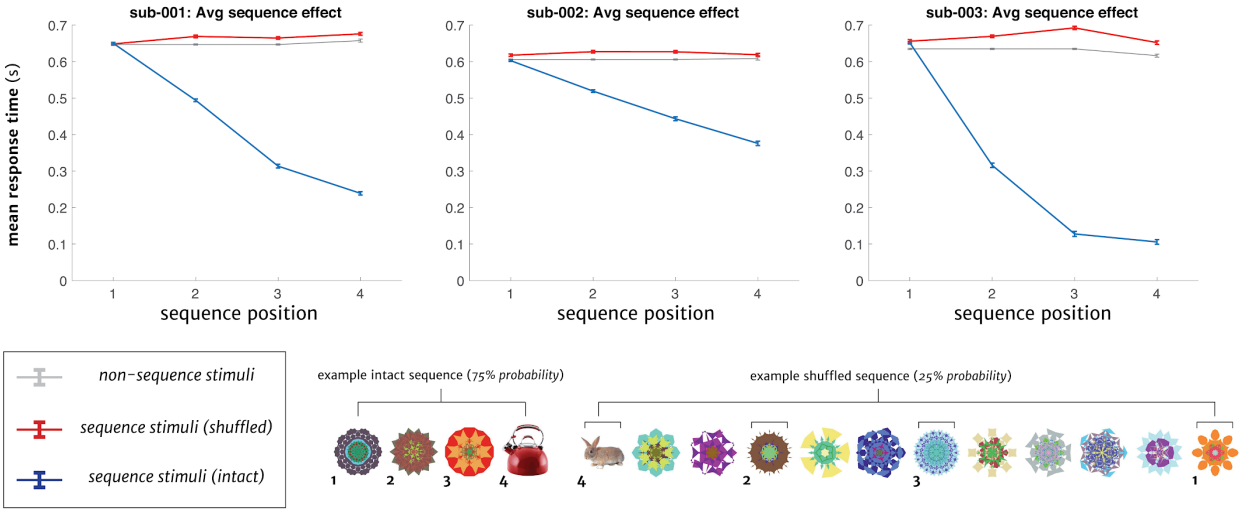
First, we constructed an *item-level* model for individual stimulus representations by comparing the on-diagonal correlations (between trials featuring the same stimulus) and off-diagonal correlations for the six trained stimuli not included in any of the learned sequences (**Figure 5c**). Second, we operationalized a category-level model by testing for an interaction in the off-diagonal correlations among all pairs of 18 trained (**Figure 5b,** dark blue) stimuli and the six novel (**Figure 5b,** light blue) stimuli within each session. Finally, we constructed two separate models to test for representations of stimulus sequences from the SRT task. The first sequence representational structure was a *within-sequence* model in which off-diagonal correlations of trained stimuli within the same sequence were compared to the correlations between stimuli in sequences to the trained stimuli not in sequences (**Figure 6a**). Next, we constructed a *between-sequence* model to test for an interaction in the similarity of stimuli between *different* sequences (**Figure 6b**,), again compared to a baseline of correlations to trained stimuli not in sequences. A final follow-up model directly tested the within versus between-sequence stimulus correlations, with no differences found across conditions. For the analysis of off-diagonal correlations among trained stimuli in **Figure 5a**, we excluded the correlations between stimulus pairs within the same sequence from the SRT task. To determine if there were changes in pattern similarity across training, we used mixed nonlinear models with the *beta* values from the toy matrix regressor ("pattern strength" values) values as the outcome variable and session number (mean-centered) as predictors. For all models, ROIs with a significant change in the pattern strength across training (significant value of the linear *b* parameter, see *Statistical methods*) are bolded in **Figure 5** and **Figure 6**. We also included early visual and lateral occipital ROIs in the pattern similarity analyses to determine what representational changes are specific to the PFC versus early and higher-order sensory areas.

Statistical methods

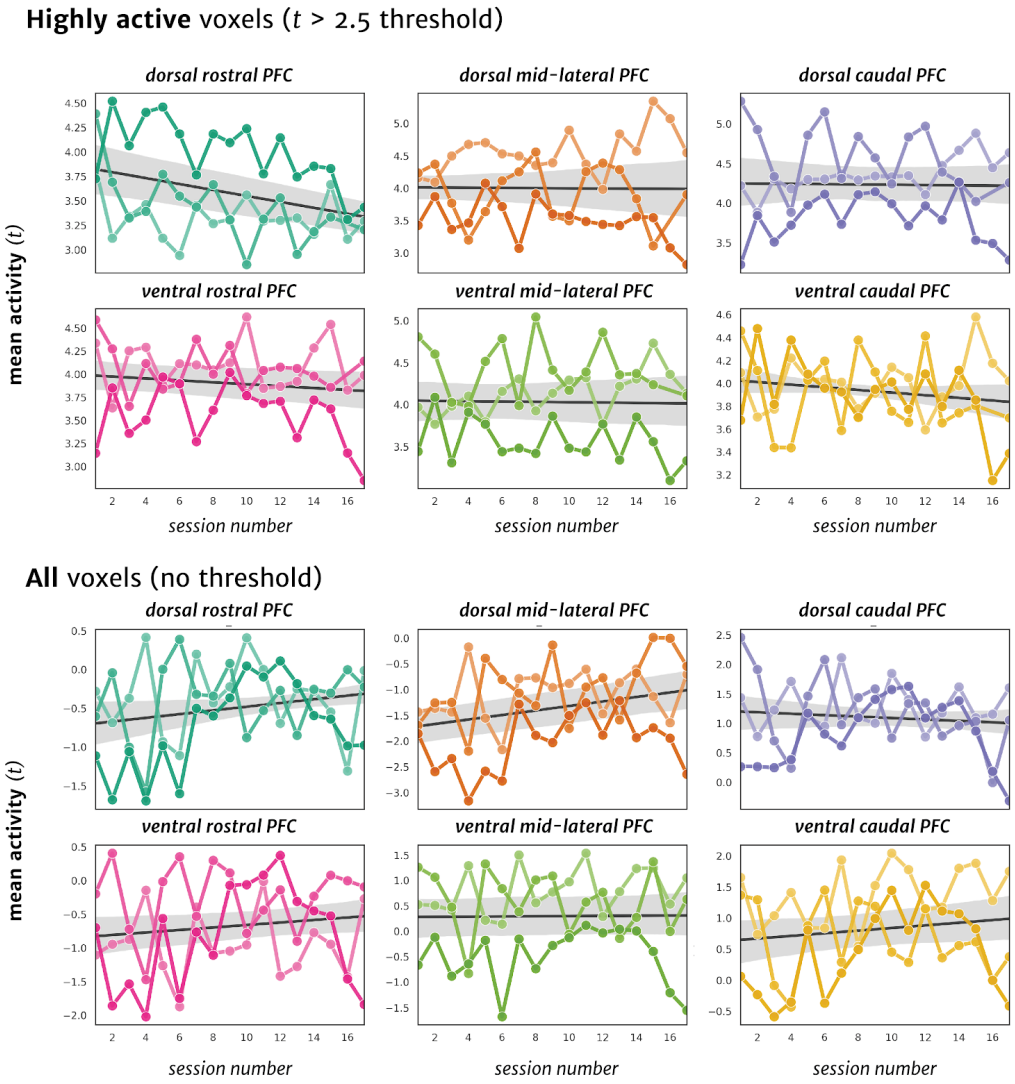
All changes across training were analyzed using mixed nonlinear models, implemented in the *nlme* library in R (https://cran.r-project.org/web/packages/nlme/index.html). For the nonlinear models, we implemented a second-order polynomial function ($y = a \cdot x^2 + b \cdot x + c$) with all three parameters (a, b, c) in the function used as both fixed and random effects (random effects: $a + b + c \sim 1 \mid subject$). The linear term of the model (b) was used to test for significance of increases or decreases in outcome variables across sessions. For all models, the session number (predictor) variable was mean-centered in order to facilitate interpretation of the direction of change of the nonlinear models ($b > 0$: increasing, $b < 0$: decreasing). Starting values for the nonlinear model fitting were obtained using the selected data averaged across conditions and groups, implemented in the *polyfit* function for R. If nonlinear models failed to converge with full random effects, the nonlinear term (*a)* was removed as a random effect and the model was run again (random effects: $b + c \sim 1 \mid subject$). For all results, changes over time and conditional interactions also replicated when using mixed linear models. Voxel-wise regression models and

selectivity *F*-test measures were also calculated using *statsmodels* (https://www.statsmodels.org/stable/index.html) and *Scipy* (https://www.scipy.org/) functions in Python. Neuroimaging files were loaded and operated on using the *Nilearn* package (https://nilearn.github.io/).

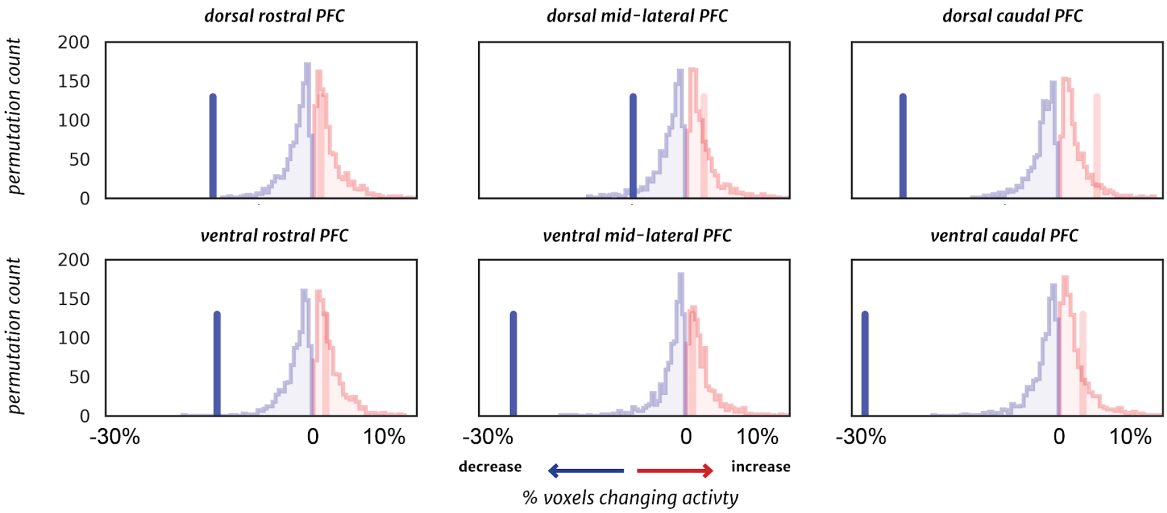
Supplemental Information



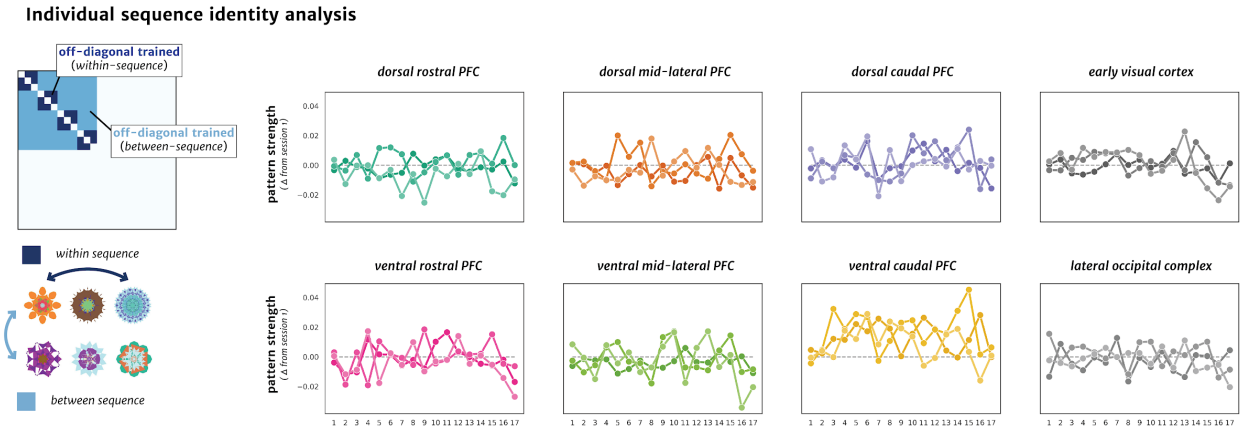
SI Figure 1. Sequence learning in the Serial Reaction Time (SRT) task.
Top: Mean response time (s) for correct trials is plotted for each participant across sequence position (1-4) for intact sequences (*blue*), compared to when the same stimuli were shown out of order (*shuffled, red*), and relative to non-sequence stimuli for reference (*gray*). All three participants showed significantly speeded responses across stimuli in intact sequences during fMRI sessions (sub-001, Position 2: $t(15) = -5.50$, $p = 6.1 \times 10^{-5}$; Position 3: $t(15) = -6.29$, $p = 1.4 \times 10^{-5}$; Position 4: $t(15) = -8.58$, $p = 3.6 \times 10^{-7;}$; sub-002, Position 2: $t(15) = -7.90$, $p = 1.0 \times 10^{-6}$; Position 3: $t(15) = -7.5$, $p = 1.8 \times 10^{-6}$; Position 4: $t(15) = -9.4$, $p = 1.1 \times 10^{-7}$; sub-003, Position 2: $t(15) = -7.80$, $p = 1.2 \times 10^{-6}$; Position 3: $t(15) = -8.6$, $p = 3.2 \times 10^{-7}$; Position 4: $t(15) = -8.6$, $p = 3.3 \times 10^{-7}$). Bottom: Examples of an intact, left, or shuffled, right, sequence in the SRT task. Intact sequences occurred with higher probability (75%) than shuffled sequences (25%). Error bars represent 68% CI (S.E.M.).



SI Figure 2. Mean WM delay activity in all lPFC regions across training.
Top: Mean activity for each fMRI session during the WM delay period for highly active voxels in each lPFC ROI, thresholded at $t > 2.5$. Specific to dorsal rostral PFC (*green*), there was a mean decrease in WM delay activity in the ROI across sessions. Bottom: For all voxels (unthresholded) in an ROI, there was only an increase in WM delay activity specific to dorsal mid-lateral PFC (*orange*).



SI Figure 3. Distribution of activity for the WM encoding epoch in PFC across the course of learning. Significant increases (*red*) or decreases (*blue*) in the percentage of voxels changing activity across training are indicated by bolded vertical lines. Null distributions were created exactly as in Figure 3, but instead using the WM encoding period activity across sessions. All ROIs show a significant proportion of voxels with a decrease in activity, with no ROIs showing an increase in WM encoding activity across training.



SI Figure 4. Analysis of individual sequence representation in WM delay activity.
(a) Left: Schematic of the model matrix for the analysis of correlations for items within the same trained sequences (dark blue, positive values) compared to correlations of items between different sequences (light blue, negative values). Right: Plots of the pattern strength across sessions for each ROI, as assessed by the model fit for the *individual sequence-level* model on the left. For visualization, all ROIs with significant changes in pattern strength across sessions are indicated with a *p*-value and bolded plot border, and pattern strength is plotted as a change from initial (session 1) baseline values. Change in pattern strength: *dorsal rostral: t(46) = -0.13, p = 0.9; dorsal mid-lateral: t(46) = -0.07, p = 0.94; dorsal caudal: t(46) = 0.55, p = 0.59; early visual: t(46) = -1.93, p = 0.06; ventral rostral: t(46) = -0.24, p = 0.8; ventral mid-lateral: t(46) = -0.78, p = 0.44; LOC: t(46) = -1.23, p = 0.22.* Each color shade represents one of the three individual participants.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*, 14.

Asp, I. E., Störmer, V. S., & Brady, T. F. (2021). Greater Visual Working Memory Capacity for Visually Matched Stimuli When They Are Perceived as Meaningful. *Journal of Cognitive Neuroscience*, *33*(5), 902–918.

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews. Neuroscience*, *10*(9), 659–669.

Badre, D., Frank, M. J., & Moore, C. I. (2015). Interactionist Neuroscience. *Neuron*, *88*(5), 855–860.

Badre, D., & Nee, D. E. (2018). Frontal Cortex and the Hierarchical Control of Behavior. *Trends in Cognitive Sciences*, *22*(2), 170–188.

Berger, M., Calapai, A., Stephan, V., Niessing, M., Burchardt, L., Gail, A., & Treue, S. (2018). Standardized automated training of rhesus monkeys for neuroscience research in their housing environment. *Journal of Neurophysiology*, *119*(3), 796–807.

Bertolero, M. A., Yeo, B. T. T., Bassett, D. S., & D'Esposito, M. (2018). A mechanistic model of connector hubs, modularity and cognition. *Nature Human Behaviour*, *2*(10), 765–777.

Beukers, A. O., Buschman, T. J., Cohen, J. D., & Norman, K. A. (2021). Is Activity Silent Working Memory Simply Episodic Memory? *Trends in Cognitive Sciences*, *25*(4), 284–293.

Bhandari, A., Gagne, C., & Badre, D. (2018). Just above Chance: Is It Harder to Decode Information from Prefrontal Cortex Hemodynamic Activity Patterns? *Journal of Cognitive Neuroscience*, *30*(10), 1473–1498.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536.

Birman, D., & Gardner, J. L. (2016). Parietal and prefrontal: categorical differences? [Review of *Parietal and prefrontal: categorical differences?*]. *Nature Neuroscience*, *19*(1), 5–7.

Blalock, L. D. (2015). Stimulus familiarity improves consolidation of visual working memory representations. *Attention, Perception & Psychophysics*, *77*(4), 1143–1158.

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(27), 7459–7464.

Buschkuehl, M., Hernandez-Garcia, L., Jaeggi, S. M., Bernard, J. A., & Jonides, J. (2014). Neural effects of short-term training on working memory. *Cognitive, Affective & Behavioral Neuroscience*, *14*(1), 147–160.

Buschkuehl, M., Jaeggi, S. M., & Jonides, J. (2012). Neuronal effects following working memory training. *Developmental Cognitive Neuroscience*, *2 Suppl 1*, S167–S179.

Chatham, C. H., Frank, M. J., & Badre, D. (2014). Corticostriatal output gating during selection from working memory. *Neuron*, *81*(4), 930–942.

Chaudhuri, R., Knoblauch, K., Gariel, M. A., Kennedy, H., & Wang, X. J. (2015). A Large-Scale Circuit Mechanism for Hierarchical Dynamical Processing in the Primate Cortex. *Neuron*, *88*(2), 419–431.

Chen, H., & Naya, Y. (2020). Forward Processing of Object-Location Association from the Ventral Stream to Medial Temporal Lobe in Nonhuman Primates. *Cerebral Cortex* , *30*(3), 1260–1271.

Christophel, T. B., Hebart, M. N., & Haynes, J. D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(38), 12983–12989.

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J. D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, *21*(2), 111–124.

Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X.-L., Wang, M., & Arnsten, A. F. T. (2018). Persistent Spiking Activity Underlies Working Memory. *The Journal of Neuroscience: The*

Official Journal of the Society for Neuroscience, *38*(32), 7020–7028.
Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, *29*(3), 162–173.
Curtis, C. E., & Sprague, T. C. (2021). Persistent Activity during Working Memory from Front to Back. In *bioRxiv* (p. 2021.04.24.441274). https://doi.org/10.1101/2021.04.24.441274
Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194.
Dang, W., Jaffe, R. J., Qi, X.-L., & Constantinidis, C. (2021). Emergence of Nonlinear Mixed Selectivity in Prefrontal Cortex after Training. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *41*(35), 7420–7434.
D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142.
Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews. Neuroscience*, *2*(11), 820–829.
Duverne, S., & Koechlin, E. (2017). Rewards and Cognitive Control in the Human Prefrontal Cortex. *Cerebral Cortex* , *27*(10), 5024–5039.
Eichenbaum, H. (2017). Prefrontal-hippocampal interactions in episodic memory. *Nature Reviews. Neuroscience*. https://doi.org/10.1038/nrn.2017.74
Eriksson, J., Vogel, E. K., Lansner, A., Bergstrom, F., & Nyberg, L. (2015). Neurocognitive Architecture of Working Memory. *Neuron*, *88*(1), 33–46.
Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R., & Gorgolewski, K. J. (2018). FMRIPrep: a robust preprocessing pipeline for functional MRI. *bioRxiv*. https://doi.org/10.1101/306951
Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*, *87*(4), 893–905.
Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, *9*(2), 195–207.
Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-Resolution Intersubject Averaging and a Coordinate System for the Cortical Surface. *Human Brain Mapping*, *8*, 272–284.
Fornito, A., Arnatkevičiūtė, A., & Fulcher, B. D. (2019). Bridging the Gap between Connectome and Transcriptome. *Trends in Cognitive Sciences*, *23*(1), 34–50.
Fukuda, K., & Woodman, G. F. (2017). Visual working memory buffers information retrieved from visual long-term memory. *Proceedings of the National Academy of Sciences*, 201617874.
Funahashi, S., Bruce, C., & Goldman-Rakic, P. S. (1989). Mnemonic encoding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61*.
Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1990). Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. *Journal of Neurophysiology*, *63*(4), 814–831.
Fuster, J., & Alexander, G. (1971). Neuron activity related to short-term memory. *Science*, *173*, 652–654.
Garcia-Cabezas, M. A., Zikopoulos, B., & Barbas, H. (2019). The Structural Model: a theory linking connections, plasticity, pathology, development and evolution of the cerebral cortex. *Brain Structure & Function*. https://doi.org/10.1007/s00429-019-01841-9
Garcia, K. E., Robinson, E. C., Alexopoulos, D., Dierker, D. L., Glasser, M. F., Coalson, T. S., Ortinau, C. M., Rueckert, D., Taber, L. A., Van Essen, D. C., Rogers, C. E., Smyser, C. D., & Bayly, P. V. (2018). Dynamic patterns of cortical expansion during folding of the preterm human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(12), 3156–3161.
Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129–135.
Ghazizadeh, A., Griggs, W., Leopold, D. A., & Hikosaka, O. (2018). Temporal-prefrontal cortical network for discrimination of valuable objects in long-term memory. *Proceedings of the National*

Academy of Sciences of the United States of America, *115*(9), E2135–E2144.
Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178.
Goldman-Rakic, P. S. (1995). Cellular Basis of Working Memory. *Neuron*, *14*, 477–485.
Gorgolewski, K. J., Auer, T., Calhoun, V. D., Cameron Craddock, R., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nolan Nichols, B., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 1–9.
Gorgolewski, K. J., Esteban, O., Ellis, D. G., Notter, M. P., Ziegler, E., Johnson, H., Hamalainen, C., Yvernault, B., Burns, C., Manhães-Savio, A., Jarecka, D., Markiewicz, C. J., Salo, T., Clark, D., Waskom, M., Wong, J., Modat, M., Dewey, B. E., Clark, M. G., … Ghosh, S. (2017). *Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. 0.13.1*. https://doi.org/10.5281/zenodo.581704
Goulas, A., Uylings, H. B., & Stiers, P. (2014). Mapping the hierarchical layout of the structural network of the macaque prefrontal cortex. *Cerebral Cortex* , *24*(5), 1178–1194.
Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72.
Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.
Hoskin, A. N., Bornstein, A. M., Norman, K. A., & Cohen, J. D. (2019). Refresh my memory: Episodic memory reinstatements intrude on working memory maintenance. *Cognitive, Affective & Behavioral Neuroscience*, *19*(2), 338–354.
Ito, T., Kulkarni, K. R., Schultz, D. H., Mill, R. D., Chen, R. H., Solomyak, L. I., & Cole, M. W. (2017). Cognitive task information is transferred between brain regions via resting-state network topology. *Nature Communications*, *8*(1), 1027.
Jackson, M. C., & Raymond, J. E. (2008). Familiarity enhances visual working memory for faces. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(3), 556–568.
Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841.
Kim, H. F., Ghazizadeh, A., & Hikosaka, O. (2015). Dopamine Neurons Encoding Long-Term Memory of Object Value for Habitual Behavior. *Cell*, *163*(5), 1165–1175.
Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*(7), 317–324.
Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*(5648), 1181–1185.
Lara, A. H., & Wallis, J. D. (2015). The Role of Prefrontal Cortex in Working Memory: A Mini Review. *Frontiers in Systems Neuroscience*, *9*, 173.
LaRocque, J. J., Lewis-Peacock, J. A., & Postle, B. R. (2014). Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Human Neuroscience*, *8*.
Leavitt, M. L., Mendoza-Halliday, D., & Martinez-Trujillo, J. C. (2017). Sustained Activity Encoding Working Memories: Not Fully Distributed. *Trends in Neurosciences*, *40*(6), 328–346.
Lee, S. H., Kravitz, D. J., & Baker, C. I. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature Neuroscience*, *16*(8), 997–999.
Lewis-Peacock, J. A., & Norman, K. A. (2014). Competition between items in working memory leads to forgetting. *Nature Communications*, *5*, 5768.
Lorenc, E. S., & Sreenivasan, K. K. (2021). Reframing the debate: The distributed systems view of working memory. *Visual Cognition*, 1–9.
Mackey, W. E., Devinsky, O., Doyle, W. K., Meager, M. R., & Curtis, C. E. (2016). Human Dorsolateral Prefrontal Cortex Is Not Necessary for Spatial Working Memory. *The Journal of Neuroscience: The*

Official Journal of the Society for Neuroscience, *36*(10), 2847–2856.
Mendoza-Halliday, D., Torres, S., & Martinez-Trujillo, J. C. (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nature Neuroscience*, *17*(9), 1255–1262.
Meyer, T., Qi, X.-L., Stanford, T. R., & Constantinidis, C. (2011). Stimulus selectivity in dorsal and ventral prefrontal cortex after training in working memory tasks. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(17), 6266–6276.
Milham, M. P., Ai, L., Koo, B., Xu, T., Amiez, C., Balezeau, F., Baxter, M. G., Blezer, E. L. A., Brochier, T., Chen, A., Croxson, P. L., Damatac, C. G., Dehaene, S., Everling, S., Fair, D. A., Fleysher, L., Freiwald, W., Froudist-Walsh, S., Griffiths, T. D., … Schroeder, C. E. (2018). An Open Resource for Non-human Primate Imaging. *Neuron*, *100*(1), 61–74.e2.
Miller, E. K., & Cohen, J. D. (2001). AN INTEGRATIVE THEORY OF PREFRONTAL CORTEX FUNCTION. *Annual Review of Neuroscience*, *24*, 167–202.
Miller, E. K., & Fusi, S. (2013). Limber neurons for a nimble mind. *Neuron*, *78*(2), 211–213.
Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working Memory 2.0. *Neuron*, *100*(2), 463–475.
Miller, J. A., Voorhies, W. I., Lurie, D. J., D'Esposito, M., & Weiner, K. S. (2021). Overlooked Tertiary Sulci Serve as a Meso-Scale Link between Microstructural and Functional Properties of Human Lateral Prefrontal Cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *41*(10), 2229–2244.
Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., & Malach, R. (2005). Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science*, *309*(5736), 951–954.
Mumford, J. A., Davis, T., & Poldrack, R. A. (2014). The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage*, *103*, 130–138.
Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, *59*(3), 2636–2643.
Murray, J. D., Bernacchia, A., Roy, N. A., Constantinidis, C., Romo, R., & Wang, X. J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(2), 394–399.
Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–227.
Naya, Y., Yoshida, M., & Miyashita, Y. (2001). Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science*, *291*(5504), 661–664.
Nee, D. E., & Jonides, J. (2011). Dissociable contributions of prefrontal cortex and the hippocampus to short-term memory: evidence for a 3-state model of memory. *NeuroImage*, *54*(2), 1540–1548.
Oberauer, K. (2009). Design for a Working Memory. In *The Psychology of Learning and Motivation* (pp. 45–100).
Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, *7*(1), 75–79.
Park, S. H., Russ, B. E., McMahon, D. B. T., Koyano, K. W., Berman, R. A., & Leopold, D. A. (2017). Functional Subpopulations of Neurons in a Macaque Face Patch Revealed by Single-Unit fMRI Mapping. *Neuron*. https://doi.org/10.1016/j.neuron.2017.07.014
Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 781–795.
Petrides, M. (2019). *Atlas of the Morphology of the Human Cerebral Cortex on the Average MNI Brain* (1st ed.). Elsevier.
Qi, X. L., Riley, M. R., & Constantinidis, C. (2019). Working memory capacity is enhanced by distributed prefrontal activation and invariant temporal dynamics. *Proceedings of the*. https://www.pnas.org/content/116/14/7095.short
Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences*, *9*(8), 374–380.

Ranganath, C., Johnson, M. K., & D'Esposito, M. (2003). Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia*, *41*(3), 378–389.

Riggall, A. C., & Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(38), 12990–12998.

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*(7451), 585–590.

Riley, M. R., Qi, X. L., Zhou, X., & Constantinidis, C. (2018). Anterior-posterior gradient of plasticity in primate prefrontal cortex. *Nature Communications*, *9*(1), 3790.

Romo, R., Brody, C. D., Hernández, A., & Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, *399*(6735), 470–473.

Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature*, *354*(6349), 152–155.

Sarma, A., Masse, N. Y., Wang, X.-J., & Freedman, D. J. (2016). Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. *Nature Neuroscience*, *19*(1), 143–149.

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology: CB*, *22*(17), 1622–1627.

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*, 8151.

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(11), 4447–4452.

Serences, J. T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vision Research*, *128*, 53–67.

Sommer, T. (2017). The Emergence of Knowledge and How it Supports the Memory for Novel Related Information. *Cerebral Cortex* , *27*(3), 1906–1921.

Song, X., García-Saldivar, P., Kindred, N., Wang, Y., Merchant, H., Meguerditchian, A., Yang, Y., Stein, E. A., Bradberry, C. W., Ben Hamed, S., Jedema, H. P., & Poirier, C. (2021). Strengths and challenges of longitudinal non-human primate neuroimaging. *NeuroImage*, *236*, 118009.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386.

Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*(2), 82–89.

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*(2), 364–375.

Supèr, H., Spekreijse, H., & Lamme, V. A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science*, *293*(5527), 120–124.

Szczepanski, S. M., & Knight, R. T. (2014). Insights into human behavior from lesions to the prefrontal cortex. *Neuron*, *83*(5), 1002–1018.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320.

Vallentin, D., Bongard, S., & Nieder, A. (2012). Numerical rule coding in the prefrontal, premotor, and posterior parietal cortices of macaques. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(19), 6621–6630.

Wagstyl, K., Larocque, S., Cucurull, G., Lepage, C., Cohen, J. P., Bludau, S., Palomero-Gallagher, N., Lewis, L. B., Funck, T., Spitzer, H., Dickscheid, T., Fletcher, P. C., Romero, A., Zilles, K., Amunts, K., Bengio, Y., & Evans, A. C. (2020). BigBrain 3D atlas of cortical layers: Cortical and laminar thickness gradients diverge in sensory and motor cortices. *PLoS Biology*, *18*(4), e3000678.

Wallis, J. D., & Miller, E. K. (2003). From rule to response: neuronal processes in the premotor and

prefrontal cortex. *Journal of Neurophysiology*, *90*(3), 1790–1806.
Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188–200.
Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic Maps of Visual Topography in Human Cortex. *Cerebral Cortex* , *25*(10), 3911–3931.
Wang, Y., Royer, J., Park, B.-Y., de Wael, R. V., Lariviere, S., Tavakol, S., Rodriguez-Cruces, R., Paquola, C., Hong, S.-J., Margulies, D., Smallwood, J., Valk, S., Evans, A., & Bernhardt, B. C. (2021). Long-range connections mirror and link microarchitectural and cognitive hierarchies in the human brain. In *bioRxiv* (p. 2021.10.25.465692). https://doi.org/10.1101/2021.10.25.465692
Warrington, E. K., & Shallice, T. (1969). The selective impairment of auditory verbal short-term memory. *Brain: A Journal of Neurology*, *92*(4), 885–896.
Wasmuht, D. F., Spaak, E., Buschman, T. J., Miller, E. K., & Stokes, M. G. (2018). Intrinsic neuronal dynamics predict distinct functional roles during working memory. *Nature Communications*, *9*(1), 3499.
Wickelgren, W. A. (1969). Sparing of short-term memory in an amnesic patient: implications for strength theory of memory. In *Neurocase* (Vol. 2, Issue 4, p. 259as – 298). https://doi.org/10.1093/neucas/2.4.259-as
Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society: JINS*, *17*(5), 766–780.
Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science*, *300*(5625), 1578–1581.
Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011). Adaptive coding of task-relevant information in human frontoparietal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(41), 14592–14599.
Xie, W., & Zhang, W. (2017). Familiarity increases the number of remembered Pokémon in visual short-term memory. *Memory & Cognition*, *45*(4), 677–689.
Zeithamova, D., de Araujo Sanchez, M. A., & Adke, A. (2017). Trial timing and pattern-information analyses of fMRI data. *NeuroImage*, *153*, 221–231.
Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57.

Chapter 3: Cortico-striatal output gating of working memory content

"... the neural systems that mediate the sensorimotor behavior of our ancient ancestors may have provided the foundations for modern cognitive abilities, and their consideration may shed light on the neural mechanisms that underlie human thought."

- Cisek & Kalaska (2010, *p. 289*)

This chapter contains previously published material from the following work, and permissions have been obtained from all co-authors for inclusion in this document. Supplemental information can also be found at the following reference:

Miller, J. A., Kiyonaga, A., Ivry, R. B., & D'Esposito, M. (2020, October 22). Prioritized Verbal Working Memory Content Biases Ongoing Action. *Journal of Experimental Psychology: Human Perception and Performance*. http://dx.doi.org/10.1037/xhp0000868

Abstract

Working memory (WM) holds information temporarily in mind, imparting the ability to guide behavior based on internal goals rather than external stimuli. However, humans often maintain WM content for a future task while performing more immediate actions. Consequently, transient WM representations may inadvertently influence ongoing (but unrelated) motor behavior. Here, we tested the impact of WM on adult human action execution and examined how the attentional or "activation" state of WM content modulates that impact. In three dual-task experiments, verbal WM for directional words influenced the trajectory and speed of hand movements performed during WM maintenance. This movement bias was also modulated by the attentional state of the WM content. Prioritized WM content strongly influenced actions during WM maintenance, while de-prioritized WM content was less influential. In sum, WM can unintentionally shape ongoing motor behavior, but the behavioral relevance of WM content determines the degree of influence on motor output.

Significance statement

Working memory allows us to keep information actively in mind, so that we can use that information to achieve our moment-to-moment goals. However, this working memory maintenance process may unintentionally impact our interactions with the environment, and can occasionally interfere with our immediate external goals. This study formalizes the everyday "action slips" that humans commit when we type out or say the wrong word aloud in conversation because it was held in mind for a different goal. The results show that internally maintained content can influence the direction and speed of hand movements that are executed during working memory maintenance. However, the extent of this action interference varies with the relevance of the maintained content to either immediate or temporally-extended task goals. That is, working memory can bias our actions, but we can control the behavioral state of working memory content to reduce the likelihood of these everyday errors.

Introduction

As humans engage in complex cognition, our thoughts can inadvertently influence our interactions with the environment. Imagine, for instance, accidentally typing out or saying the wrong word aloud in conversation because it was currently on your mind. Here, we test the idea that such everyday cognitive slips emerge from the typically adaptive processes by which working memory (WM) guides behavior. WM maintains temporary mnemonic representations that can support perceptual continuity and attentional orienting, but a core function of WM is also to steer goal-directed actions (Fuster & Alexander, 1971; van Ede, Chekroud, Stokes, & Nobre, 2019). Recent studies have begun to highlight the influential role of WM in preparing volitional motor responses (Belopolsky & Theeuwes, 2011; Zokaei, Board, Manohar, & Nobre, 2019), as well as the reciprocal role of motor systems in supporting WM maintenance (Hanning, Jonikaitis, Deubel, & Szinte, 2016; Ohl & Rolfs, 2017). While WM is classically described as a system for short-term information storage, some theories assert that it would be better construed as intention to perform an action (Fuster, 1990, 2004, 2015; Postle, 2006; Theeuwes, Belopolsky, & Olivers, 2009). If that is the case, then WM content may bias ongoing motor behavior, even when the content is not directly task relevant. The current study tests the boundaries of the linkage between WM and action control. We investigate whether WM impacts action execution during maintenance, and how that impact is modulated by task goals.

WM maintenance can bias visual attention toward related content in the environment, even at the expense of the current task (Soto, Hodsoll, Rotshtein, & Humphreys, 2008). However, this influence of WM can be strategically modulated (Carlisle & Woodman, 2011; Kiyonaga, Egner, & Soto, 2012). For instance, when one among several WM items is flagged as relevant (i.e., 'retro-cued'), attended WM items evoke more detectable neural traces with fMRI or M/EEG (LaRocque, Lewis-Peacock, & Postle, 2014; Wolff, Jochim, Akyurek, & Stokes, 2017), and such prioritized content biases visual attention more strongly (Mallett & Lewis-Peacock, 2018; Olivers, Peters, Houtkamp, & Roelfsema, 2011; van Moorselaar, Theeuwes, & Olivers, 2014). These findings align with a biased competition account of visual attention, wherein visual representations compete for attention in a weighted manner (Desimone & Duncan, 1995). This model further predicts that WM should have a similar influence on competing response systems. That is, if actively maintained WM content biases perceptual representations that guide attention, it may also bias motor preparation representations that guide action (Meiran, Cole, & Braver, 2012; Oberauer, 2010; Theeuwes et al., 2009). Yet we often maintain WM content that is relevant for a future goal while engaged in more immediate actions. As a result, actively maintained WM representations may bias ongoing (but unrelated) motor behavior.

While prioritized WM content is considered to be in a privileged, active state (Zokaei, Manohar, Husain, & Feredoes, 2014), deprioritized content is relatively less immediately relevant and is considered "unattended" (LaRocque et al., 2014; Lewis-Peacock, Drysdale, Oberauer, & Postle, 2011). It may be maintained with lower activity levels (Bays & Taylor, 2018) or in a distinct latent format (Rose et al., 2016; Sprague, Ester, & Serences, 2016), but the cognitive, neural, and behavioral status of this deprioritized content is currently unclear (Mallett & Lewis-Peacock, 2018; Manohar, Zokaei, Fallon, Vogels, & Husain, 2019; Nobre & Stokes, 2019; Park, Sy, Hong, & Tong, 2017). If deprioritized content is maintained at a quantitatively lower level of activation, it

may exert a diminished but still measurable influence on behavior. If, on the other hand, it is maintained in a qualitatively distinct state, it may be prevented from spilling over into behavior. Internally oriented attention processes modulate visual WM performance (Souza & Oberauer, 2016), and are theorized to modulate the nature of visual WM representations (Stokes, 2015; van Loon, Olmos-Solis, Fahrenfort, & Olivers, 2018; Wolff, Ding, Myers, & Stokes, 2015). Here, we test the idea that such attentional prioritization processes should also determine how WM information influences ongoing motor behavior. In other words, we ask whether relative WM priority levels drive (sometimes erroneous) actions that occur during maintenance. We therefore employ an intervening motor task as a behavioral "probe" into the activation state of the WM content.

In theories of motor function, goal-potentiated frontal cortical representations feed into a basal ganglia gating mechanism whereby only the most active representations surpass the threshold to drive actions (Ivry & Spencer, 2004). The behavioral influence of WM has been theoretically attributed to a similar output gating function (Wallis, Stokes, Cousijn, Woolrich, & Nobre, 2015). In this model, attentional selection transforms WM representations from suspended internal maintenance into a behavior-driving state (Myers, Stokes, & Nobre, 2017). The presumed gating system also tracks the utility of WM representations, defining which ones should be selected (Chatham & Badre, 2015; Cools, Ivry, & D'Esposito, 2006). If the activity state of WM content determines such a gating process, task-irrelevant representations could be gated out when they are activated above threshold. While output gating has typically been considered to result from volitional selection of WM content, it could theoretically sometimes drive incorrect actions. However, in a complex task, multiple rules must be tracked and segregated for successful behavior, and this may be accomplished through hierarchical control functions which regulate gating behavior (Badre & Nee, 2018). Here, we also examine how item-level attentional selection (which may promote output gating) interacts with task-level goal maintenance functions to control WM-guided behavior.

Patients with frontal lesions sometimes display contextual *action slips,* like sprinkling salt into tea instead of on food (Schwartz, 1995). These deficits may be an exaggerated form of the everyday slips humans commit when thought content unintentionally infiltrates behavior. But what determines which representations surpass the action threshold? And how might an adaptive WM process interact with contextual task demands? To experimentally formalize action slips, we created scenarios wherein hand movements were executed during WM maintenance. In this dual-task setting, cued movement directions could be either compatible or incompatible with the meaning of the maintained content. We manipulated the predictive relationship between WM and motor goals across a task context (i.e., varying the proportion of compatible trials), as well the attentional priority level of WM content for trial-by-trial WM goals (i.e., retro-cueing). Across three experiments, this study examines whether WM maintenance biases intervening motor action, and if so, how task context and behavioral relevance influence this bias.

Experiment 1: Block-wise WM-relevance manipulation

Methods

Participants. Participants were recruited from the Berkeley community, gave informed consent in accordance with the University of California Berkeley Institutional Review Board (IRB), and received either course credit or $20 per hour for participation. We aimed to recruit 30 participants for each experiment. This sample size was estimated from previous experiments that used a WM dual-task structure and a similar statistical comparison (e.g., congruent vs. incongruent Stroop trials; Kiyonaga & Egner, 2014). We expected a similar effect size, around Cohen's d = 0.7, for our primary effect of WM compatibility. With alpha = 0.05, this would yield a power of 0.96. Individuals were excluded if accuracy was below 60% or responses were entered on fewer than ⅔ trials for either the motor or WM tasks. Experiment 1 was administered to 32 participants, but 3 were excluded for failing to meet the accuracy threshold. Therefore, analyses included 29 participants (9 male; mean age = 20.0 y, range = 18-24).

Task overview. The goal was to simulate an everyday situation where information is maintained for future use while performing immediate actions. We therefore interleaved a verbal delayed recognition test with a simple motor task (**Fig. 1a**). On each trial, participants were instructed to remember a directional word ('up,' 'down,' 'left,' or 'right'). Then, during the WM delay, they were visually cued to move the mouse and click on a target located at one of four cardinal screen positions (top, bottom, left, or right). After the motor task, participants were tested on their memory for the sample word. The meaning of the verbal WM content could be either compatible (e.g., remember 'left', click inside leftward box) or incompatible (e.g., remember 'left', click inside rightward box) with the direction of the cued hand movement. The task therefore required maintenance of multiple rules and representations for WM and motor components, which were sometimes in conflict with each other.

To examine how the priority level of WM representations might modulate their influence over behavior, we developed three variations of this basic task. Experiment 1 manipulated the relative value of the WM content across block conditions by varying the predictive utility of WM to the motor task. That is, the ratio of compatible to incompatible trials was varied across three task block contexts. Experiment 2 manipulated the trial-by-trial priority status of individual WM items to the WM test—with retro-cues to shift attention among simultaneously remembered items—while keeping the WM relationship to the motor task unchanged. Experiment 3 combined elements of Experiments 1 and 2, to examine the contributions of modulating WM at the level of task goals vs. item representations. All data and code are available on the Open Science Framework: https://bit.ly/2RlYRm5

Stimuli and procedure. All tasks were programmed using Psychtoolbox functions (Brainard, 1997) (http://psychtoolbox.org/) in Matlab (https://www.mathworks.com/), along with custom scripts to track mouse positions. Participants sat ~60 cm from a 23 in. screen. The WM stimuli consisted of directional words ('up,' 'down,' 'left,' or 'right'), presented in black (visual angle ~ 1.2º) on a neutral grey background (RGB: [128,128,128]). Every trial began with a 2 sec intertrial interval (ITI). Then a WM sample word appeared centrally for 1 sec. After a total delay of 5 sec, a WM probe word (selected from the same set as the WM samples) appeared centrally underneath a

question mark. The WM task was to make a keyboard button press indicating whether the probe word was a match ('S' key) or non-match ('D' key) to the WM sample. Match and non-match WM probes were equally likely (50% match / 50% non-match) in all experiments.

During the WM delay, participants completed a manual action task. A central filled colored square (i.e., the cue) was flanked by unfilled square boxes (i.e., the targets) at each of four locations: to the top, bottom, left, and right of center. The central square could be one of four colors (RGB: green = [122,164,86], pink = [198,89,153], orange = [201,109,68], blue = [119,122,205]), which were chosen to be maximally distinct, matched on saturation and brightness, and color-blind friendly (http://tools.medialab.sciences-po.fr/iwanthue/). Each color was instructed to cue one of four screen locations: green = *left,* pink = *right,* orange = *up,* blue = *down*. The target boxes were equidistant from the central color cue and each other (size ~ 3.7º, distance from center ~ 9.3º). The motor task was to move the mouse and click inside the target box at the location cued by the color. The motor task therefore required a symbolic transformation from color to location, which was meant to engage the goal representation circuitry involved in gating motor behaviors (O'Reilly & Frank, 2005; Oliveira & Ivry, 2008). The motor task epoch ended when a cursor click was recorded in any of the target locations, or when a 2 sec response deadline passed.

The sequence of one complete dual-task trial started with a 2 sec ITI, followed by a 1 sec WM sample display, then a 2 sec fixation delay. After this first delay, the motor task display appeared for 2 sec, followed by another fixation delay of 1 sec, and then finally the WM probe display for 2 sec (**Fig. 1a,** left). There were two primary trial types: compatible trials, wherein the meaning of the WM word matched the cued direction of movement, and incompatible trials, wherein the WM word was paired with any of the three non-matching movement cues. The ratio of compatible to incompatible trials was manipulated across a given task block. Blocks contained either 80%, 50%, or 20% compatible trials (**Fig. 1a,** right). In "high compatibility" blocks (80% compatible), the WM sample meaning usually helped the motor task, as it corresponded to the directional goal of the upcoming movement. In "middle compatibility" blocks (50% compatible), the WM content was equally likely to be helpful or harmful to the motor task on any given trial. In "low compatibility" blocks (20% compatible) the WM sample meaning usually differed from the motor task target, and was therefore unhelpful. To minimize probabilistic learning effects, participants were explicitly informed about the percentage of compatible trials at the start of each block.

In order to learn the color-direction response mapping, participants practiced at least 12 trials of the motor task before each experiment. Then participants completed one 6-trial practice block of each condition (with feedback for motor and WM response accuracy) before completing three 30-trial experimental blocks of each condition (without feedback; 9 blocks total). Participants therefore completed 90 trials in each block condition and 135 trials of each compatibility condition across blocks (72 incompatible/18 compatible trials across "low compatibility" blocks, 45 incompatible/45 compatible trials across "middle compatibility" blocks, and 18 incompatible/72 compatible trials across "high compatibility" blocks). The first block was always middle compatibility (50% compatible), while the predictability conditions occurred in random order for the remaining blocks. The difference in motor behavior on compatible vs. incompatible trials—or the 'compatibility effect'—will serve here as an operational index of the influence of WM over ongoing action.

Movement trajectory analysis. Mouse positioning data was tracked across the motor task to assess the influence of WM content on the direction of hand movements. To define when hand movement trajectories were curved away from the target location, we created a circle around the start position with a radius of ¼ the distance to the target. Trajectories were considered precise if they first crossed that boundary within 45° of the correct response axis, but were classified as *course adjustments* if they crossed that boundary at a wider angle than 45° before terminating at the correct target (**Fig. 1b**). All cursor trajectories were rotated to a common axis for comparison. Because the number of compatible and incompatible trials varied across block conditions in Experiment 1, we calculated the proportion of corrected movements as the number of *course adjustments* divided by the total trial number of that type. Finally, to test if course adjustment trajectories were biased specifically toward the direction held in WM, we analyzed trajectory data for each incompatible trial and categorized whether the exit angle of the initial movement matched the direction held in WM. We tested the proportion of trials matching the WM direction against 33.3%—the probability of moving randomly into one of the non-target directions on an incompatible trial. This is a conservative comparison, as the chance of randomly moving toward any of the four possible movement targets would be 25%.

Movement speed measures: Movement initiation—also sometimes referred to as 'reaction time'—was defined as the time from the onset of the color cue until the cursor first crossed a radius of 30 pixels from the starting position. *Movement duration*—also sometimes referred to as 'movement time'—was defined as the amount of time after movement initiation until a click was made within one of the movement targets.

Quality control criteria. Trials were excluded if no WM probe response was made. For analyses of motor response speeds, outlier trials were excluded if a measurement was greater than 3 standard deviations away from the participants' mean, or if the motor task response was inaccurate. In Experiment 1, 2.8% of total trials were excluded as response speed outliers, 1.2% as nonresponse trials, and 2.7% as response errors.

Analysis strategy: For all measures, we conducted a 2 (*trial compatibility*: compatible vs. incompatible) × 3 (*block predictability*: low vs. middle vs. high compatibility) repeated measures ANOVA. To decompose any significant interactions, we conducted one-way ANOVAs of block predictability separately for compatible and incompatible trials. All ANOVA main effects and interactions are reported with the generalized-eta-squared ( ) effect size measure. This estimate indexes the proportion of variability in the outcome measure associated with a given variable and generalizes across within- and between-subjects designs (Fritz, Morris, & Richler, 2012). All post-hoc t-tests are reported with the Cohen's d effect size and a bootstrapped (n = 10,000 bootstraps) 95% confidence interval for the Cohen's d value.

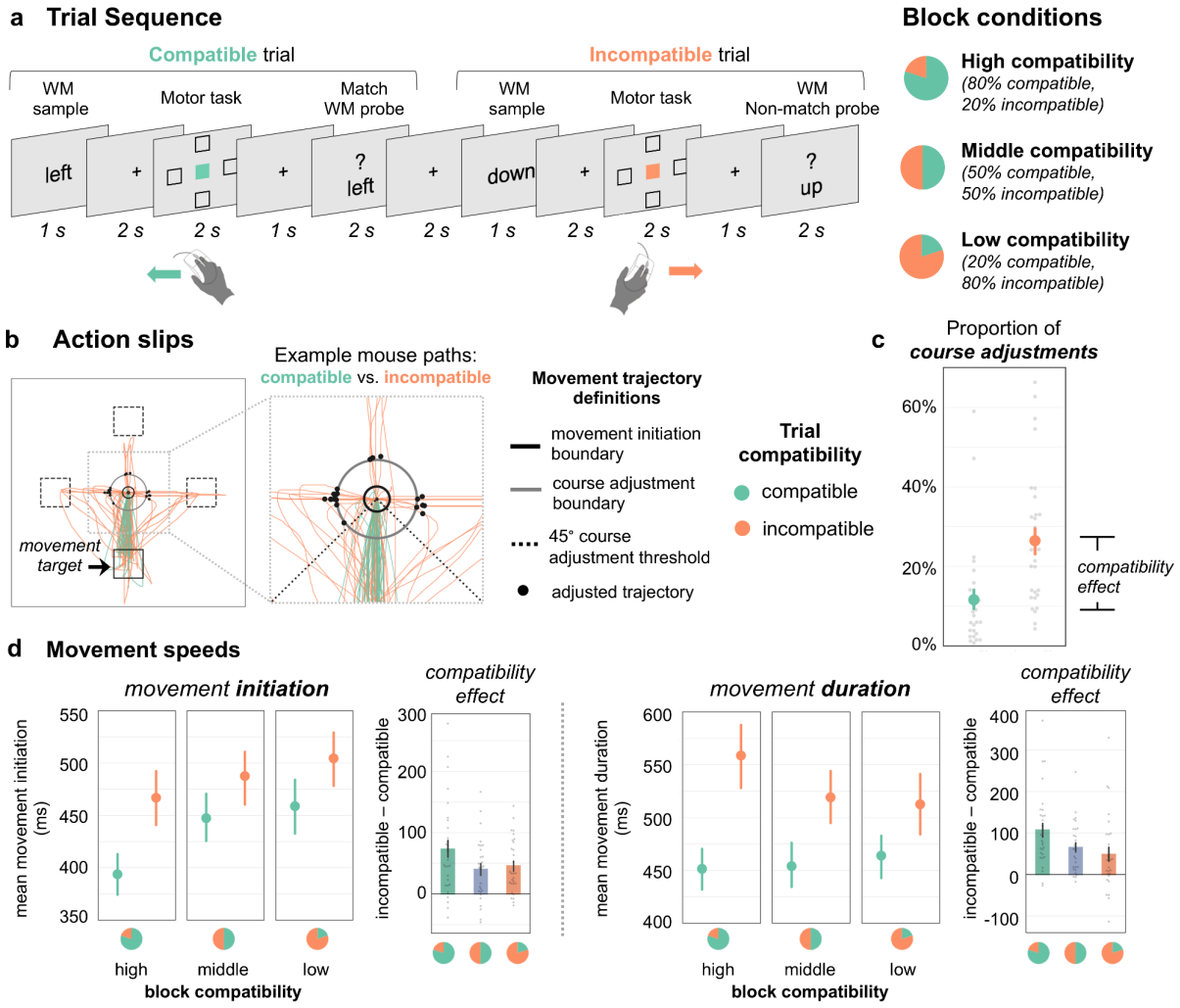


Fig 1. Experiment 1 task design and results. (a) Example *compatible* and *incompatible* trial sequences (left), which were delivered in 3 block conditions (right). **(b)** Movement trajectories for *compatible* (green) and *incompatible* trials (orange) from an example subject on middle compatibility blocks. Detail illustrates criteria for categorizing trials as "course adjustments." **(c)** The proportion of course adjustments on *compatible* (green) and *incompatible* (orange) trials, collapsed across block condition. **(d)** Point plots show mean movement speeds for all trial and block conditions, while barplots show the difference between *compatible* (green) and *incompatible* (orange) for each block condition. Left: *Movement initiation*, or 'reaction time'. Right: *Movement duration*, or 'movement time'. Error bars represent SEM. Gray dots are data points from individual participants.

Results

WM accuracy. For all experiments, WM probe accuracy was > 90% correct, confirming that participants completed the task as instructed. Neither trial compatibility nor block conditions significantly modulated WM probe accuracy (**SI Exp. 1 Results**). As our goal was to assess the influence of WM over motor behavior, the remaining analyses examine motor task performance.

Movement accuracy and action slips. We first examined whether WM content impacted the direction of cued hand movements. Mouse clicking accuracy (% correct of click location) was reliably worse on incompatible (96%) vs. compatible trials (98.5%), $F(1,28) = 26.3$, $p < .001$, $= .08$. When the meaning of the WM content was incompatible with motor goals, participants entered more responses at the wrong target location. However, there was neither a main effect of block predictability, $F(2,56) = 0.94$, $p = .47$, $= .006$, nor an interaction between factors, $F(2,56) = 0.77$, $p = .40$, $= .003$.

Movement landing positions (i.e., final click location) were highly accurate overall (97%), but the shape of movement trajectories may also be biased by WM content. For instance, incompatible trials could result in curved paths that are skewed toward the direction that matches what is held in WM. Mouse movement tracking can therefore provide a more sensitive probe into the decision processes and action execution that are influenced by ongoing cognition. Indeed, the proportion of movement *course adjustments* was ~15% greater on incompatible versus compatible trials, $F(1,28) = 45.0$, $p < .001$, $= .18$ (**Fig. 1c**). The direction of initial movement on these course adjustment trials was also most likely to match the meaning of the word held in WM, $t(28) = 3.1$, $p = .005$, $d = 0.57$, rather than being driven by general conflict processes that might impair behavior overall (**Fig. S1**). However, while this compatibility effect was descriptively largest in high compatibility blocks and smallest in low compatibility blocks, there was neither a main effect of block predictability, $F(2,56) = 0.3$, $p = .72$, $= .0006$, nor an interaction between factors, $F(2,56) = 2.4$, $p = .10$, $= .006$. In sum, action execution was biased by the meaning of WM content, but was insensitive to the block WM compatibility context. When the meaning of the WM content was incompatible with motor goals, it produced circuitous hand movement trajectories in the WM-matching direction.

Movement speeds. Multiple subprocesses may be influenced by the compatibility of the current WM content, or its relevance to the motor task. We therefore examined distinct measures of movement *initiation* and *duration*. A main effect of trial compatibility, $F(1,28) = 42.8$, $p < .001$, $= .04$, indicated that movements were initiated more slowly when the cued movement was incompatible with the WM sample (53 ms difference). A main effect of block predictability, $F(2,56) = 20.1$, $p < .001$, $= .03$, indicated that movements were also initiated more slowly overall in contexts when WM content was less likely to help motor performance. Moreover, there was an interaction between compatibility and predictability factors, $F(2,56) = 3.88$, $p = .026$, $= .003$. Separate follow-up ANOVAs for both trial types revealed effects of block predictability (compatible: $F(2,56) = 18.1$, $p < .001$, $= .05$; incompatible: $F(2,56) = 8.21$, $p < .001$, $= .01$). However, paired comparisons to the middle compatibility condition indicated that compatible movements were initiated faster in high compatibility blocks, $t(28) = 5.6$, $p < .001$, $d = 0.46$, $CI_{95\%}$ [0.25, 0.79], but no different in low compatibility blocks, $t(28) = 1.0$, $p = .32$, $d = 0.09$, $CI_{95\%}$ [-0.09, 0.29]. While incompatible trials were also initiated faster in high (vs. middle) compatibility blocks, $t(28) = 2.1$,

p = .04, d = 0.15, $CI_{95\%}$ [0.01, 0.31], they were initiated more slowly in low compatibility blocks, $t(28)$ = 2.47, p = .02, d = 0.12, $CI_{95\%}$ [0.03, 0.26]. These differences translated into a compatibility effect (i.e., incompatible - compatible RT) that was significantly greater in high compatibility blocks than in low compatibility blocks, $t(28)$ = 2.76, p = .01, d = 0.52, $CI_{95\%}$ [0.20, 0.85], (**Fig. 1d,** left).

Movement duration also showed main effects of trial compatibility, $F(1,28)$ = 33.7, p < .001, = .07, and block predictability, $F(2,56)$ = 5.01, p = .010, = .004. However, while movement *initiation* was overall speeded by a higher frequency of compatible trials, *movement duration* was slowed instead. There was also an interaction between trial and block factors, $F(2,56)$ = 8.44, p < .001, = .008, where follow-up ANOVAs revealed an effect of block predictability only for incompatible trials, $F(2,56)$ = 11.3, p = .001, = .02 (but not compatible, $F(2,56)$ = 1.1, p = .35, = .002). Incompatible movement times were relatively longer in high compatibility blocks (vs. middle compatibility), $t(28)$ = 3.83, p < .001, d = 0.26, $CI_{95\%}$ [0.12, 0.41], but low compatibility blocks were unaffected, $t(28)$ = 0.7, p = .49, d = 0.04, $CI_{95\%}$ [-0.10, 0.16]. Like *movement initiation*, the compatibility effect was greater in the high compatibility block condition compared to middle compatibility blocks, $t(28)$ = 3.82, p < .001, d = 0.56, $CI_{95\%}$ [0.29, 0.82] (**Fig. 1d,** right). Whereas the effect in *movement initiation* was driven by speeding on compatible trials, this effect in *movement duration* was driven by slowing on incompatible trials. That is, the influence of WM on movement speeds was strongest when WM content was most relevant to the motor task, although trial compatibility affected *movement initiation* and *duration* times differently. Even in low compatibility blocks alone, however, there was still a robust compatibility effect for both *movement initiation*, $t(28)$ = 5.56, p < .001, d = 0.33, $CI_{95\%}$ [0.20, 0.50], and *movement duration*, $t(28)$ = 2.87, p = .008, d = 0.35, $CI_{95\%}$ [0.12, 0.65]. To examine whether this effect might stem from a strategic carryover from high compatibility blocks when WM content was helpful, we analyzed the compatibility effect only for low compatibility (20%) blocks that were administered *before* any high compatibility (80%) blocks. The *movement initiation* benefit was present even when participants had not yet experienced any high compatibility (80%) blocks, $t(12)$ = -2.85, p = .014, d = 0.26, $CI_{95\%}$ [0.08, 0.52]. Therefore, WM content biased motor behavior before it would have been reinforced as useful for the manual task.

Finally, to address whether incompatible WM content truly slowed the rate of movement execution (rather than requiring time to redirect initially deviant movement paths), we analyzed response speeds after excluding *course adjustment* trials (**SI Exp. 1 Results**). Even precise movements were overall slowed by incompatible WM content, but the predictability between the WM content and the motor task only modulated *movement initiation* (not *duration*). In other words, the extended incompatible *movement durations* in high compatibility blocks may be explained by the longer movement paths on course adjustment trials.

Experiment 1 Discussion

Motor behavior was influenced by WM content: movement accuracy, precision, initiation, and duration were all worse when remembered words were incompatible with the intended direction of movement. Even when WM content was irrelevant to the immediate task (i.e., low compatibility blocks), it still sometimes translated into motor output. These findings mirror the attentional biasing effects of visual WM content (Desimone & Duncan, 1995; Soto et al., 2008),

suggesting that this visual cognitive framework may also be applied to understand how verbal WM content interacts with ongoing demands, and moreover, how WM content biases overt manual actions. Recent theoretical work has further described a strong functional link between visual WM and planned actions (van Ede, 2020). The current findings empirically support the notion that WM may be best understood as intention to perform an action (Theeuwes et al., 2009), extrapolating from this framework to show that the linkage may promote some unplanned influences of WM as well. Such an incidental influence of WM content has previously been labeled as "automatic," because it occurs even when the content is detrimental to ongoing processing (e.g., Soto et al., 2008). However, it could instead stem from a strategic tendency to apply WM content toward current behavior because it is typically relevant to immediate goals and would be generally adaptive to do so.

Indeed, movement speeds revealed that task context can adaptively modulate this WM influence over motor behavior. Movements were initiated faster when WM was likely to predict movement direction, but slower when it was unlikely to help. Block-level WM utility to the motor task may have lowered the decision threshold to trigger a movement, as if the WM content were gated into an action-facilitating state. However, this facilitation ultimately produced more time-consuming movement paths when WM turned out to be incompatible—i.e., slower *movement duration* in high compatibility blocks—as the motion was triggered toward the wrong location and required course adjustment to reach the target.

Here, WM content is in competition for selection with motor rules, and the likelihood for selectively gating out the WM content (rather than the correct motor rule) might theoretically increase when WM is more likely to aid motor performance (Badre, 2012). However, theories of hierarchical control also predict varying levels of segregation between representations for multiple concurrent task rules (Verbruggen, McLaren, Pereg, & Meiran, 2018). Here, when the context dictated that WM would likely help, it could have increased the relative weighting of the WM rule, making it more difficult to segregate from motor goals. This would facilitate motor behavior when the two are compatible but promote interference when incompatible, like we observed here. That is, these results may reflect control limitations in a task with competing nested rules and demands (Braem et al., 2019), rather than a modulation of gating thresholds, per se.

Experiment 2: Probabilistic retrocue manipulation

Methods

Rationale. Experiment 1 showed that the effect of WM content on motor behavior is modulated by its predictive relationship to movement goals. However, if this influence of WM does indeed stem from the activation state of the WM content—rather than simply the likelihood that the WM content will be useful—then prioritized WM content should bias actions, even when there is no relationship between WM and motor task components. Therefore, in Experiment 2, we manipulated the priority level of individual WM representations among competing alternatives. Two WM samples were presented, and trial-by-trial retrocues indicated which sample was most likely to be probed. Across all blocks of the experiment, compatible and incompatible trials were equally likely. Whereas Experiment 1 may have modulated the tonic weighting between higher order WM vs. motor task goals (i.e., compatible trials are more likely, therefore WM > motor),

explicit cueing in Experiment 2 should instead acutely modulate the weightings between concurrently maintained WM stimuli (i.e., 'left' is more likely to be tested, therefore 'left' > 'right'). If the activation state of WM content determines its degree of influence on behavior, then a prioritized (i.e., retrocued) WM item should influence ongoing actions more than a de-prioritized (i.e., uncued) item, even when both are equally likely to aid motor task performance.

Participants. A new set of 31 participants was recruited using the same guidelines and IRB approval as in Experiment 1. 28 participants were included in Experiment 2 analyses (14 male; mean age = 20.2 y, range = 18-26) after 3 exclusions for below threshold accuracy (60%).

Procedure. Experiment 2 employed the same basic dual-task structure as Experiment 1, with a few key adjustments (**Fig. 2a**). Rather than a single WM sample, two WM sample words were presented sequentially for 250 ms each, separated by a 500 ms ISI. After a 1,500 ms delay, a cue (*1*, *2*, or *X*) was displayed. A *1* or *2* served as an informative retrocue, indicating that either the first or second WM sample word was most likely to be probed. An *X* designated that the two sample words were of equal priority, as either word was equally likely to be tested (this is often referred to in the literature as a 'neutral' condition). Following a second delay of 2,000 ms, participants completed the same motor task as in Experiment 1. Before the WM probe, a probe cue appeared for 1,500 ms to indicate which item was being tested (*1* or *2*). Then the match/non-match WM probe word appeared centrally until a response, for up to 2,000 ms. Regardless of which item was retrocued earlier in the trial, the task was to compare the probe word to the corresponding WM sample: the 1st WM sample if the probe cue was a *1*, and the 2nd sample if the probe cue was a *2*. Retrocues were 90% valid, meaning that the WM sample that was cued earlier in the trial was probed on 9 out of 10 trials. Therefore, either a *1* or *2* should attentionally prioritize the cued item (and relatively de-prioritize the uncued item), whereas an *X* should result in equal priority for both items.

Three compatibility conditions were defined based on which WM item was retrocued. On *compatible* trials, the retrocued WM item was congruent with the movement direction. On *incompatible* trials, the retrocued item was incongruent with the movement direction. On *equal priority* trials, neither item was retrocued, Therefore, trials labeled *compatible* or *incompatible* only occurred when there was an informative retrocue. Retrocues only informed which WM item would be probed, but provided no information about the motor task. *Compatible*, *incompatible*, and *equal priority* trials were equally distributed in random order within each experimental block (i.e., each occurred on 1/3 of trials). Moreover, retrocue validity (i.e., whether the retrocued item was probed at the end of the trial) was unrelated to trial compatibility. In Experiment 2, therefore, the proportion of compatible trials remained constant across the experiment.

Because there were two WM sample items, they could either both be incompatible with the movement direction, or one could be compatible while the other was incompatible. *Equal priority* trials were evenly split so that one item was compatible on half of trials, while both items were incompatible on the other half of trials. For *incompatible* trials, the uncued item was selected equally often from the remaining three words, which resulted in a compatible uncued item on 1/3 of incompatible trials. However, this experiment was not optimized to examine these compatibility sub-conditions, so analyses collapse across them to maximize trial numbers in each condition (but see Experiment 3). Participants completed one 8-trial practice block (with performance feedback),

before completing at least 6 experimental blocks of 36 trials (without feedback). Each participant therefore completed at least 72 trials per condition (*compatible* / *incompatible* / *equal priority* trials) across the session.

Quality control criteria. As with Experiment 1, nonresponse trials for the WM probe, outliers for the motor response speed (> 3 s.d. from subject mean), and inaccurate motor task responses were excluded. In Experiment 2, 3.2% of total trials were excluded as response speed outliers, 1.6% as nonresponse trials, and 2.8% as response errors.

Analysis strategy. We performed one-way ANOVAs, with a factor of *cued compatibility* (*compatible* vs. *equal priority* vs. *incompatibl*e), on all measures. We then decomposed any significant effects of compatibility by calculating and comparing difference scores from the *equal priority* condition, where a 'benefit' reflects relatively faster or more precise responses, and a 'cost' reflects relatively slower or less precise responses. To determine if the influence of compatibility was reflected across each subject's distribution of response times (rather than being driven by a small number of trials on either end of the RT distribution), we analyzed the slopes of response time decile distributions (**SI Exp. 2 Methods**).

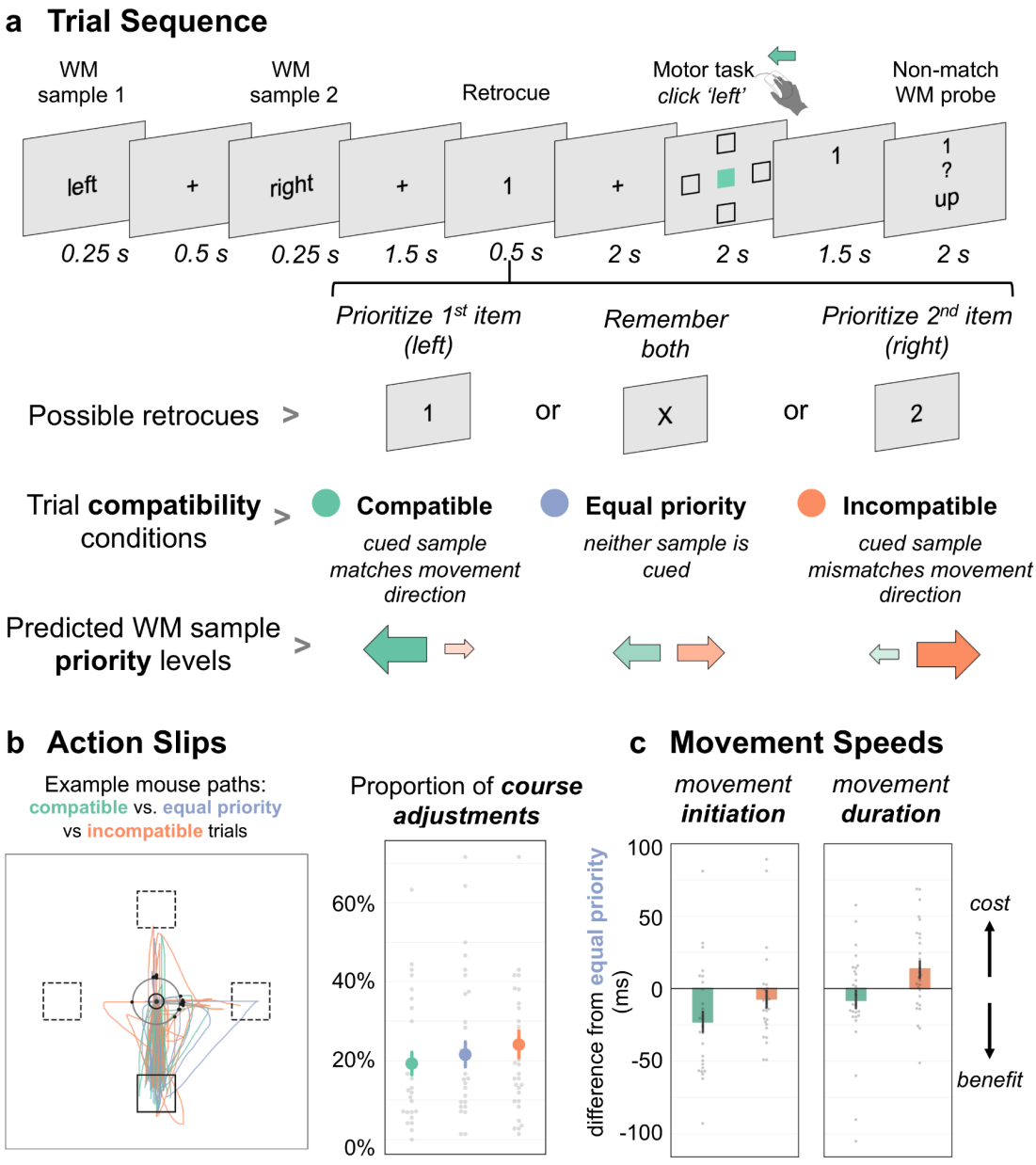


Fig 2. Experiment 2 task design and results. (a) An example trial sequence with possible retrocue and compatibility conditions. Green and orange arrows illustrate the hypothetical priority levels of the WM samples in this example trial. **(b)** Movement trajectories for *compatible* (green), *equal priority* (blue), and *incompatible* (orange) trials from an example subject on 3 blocks (left). Proportion of course adjustments on *compatible, equal priority, and incompatible* trials (right). **(c)** Benefits (green) and costs (orange) of compatible and incompatible cueing (compared to equal priority) on movement initiation (left) and duration (right) times. Error bars represent SEM. Gray dots are data points from individual participants.

Results

WM accuracy. WM probe accuracy was high (~92% correct) and was unaffected by *compatibility*, $F(2,54) = 0.4$, $p = .67$, $= .002$. To ensure that participants used the retrocue as expected, we examined the effect of retrocue validity on WM probe performance. Accuracy was better when participants were validly probed on memory for the cued item (94.7%), compared to when they were invalidly probed on memory for the uncued item (85.4%), $t(27) = 4.1$, $p < .001$, $d = 0.98$, $CI_{95\%}$ [0.61, 1.36]. Therefore, participants prioritized the cued item as expected but still remembered the uncued item well above chance.

Movement accuracy and action slips. Overall movement accuracy was high (~97%) and marginally influenced by *cued compatibility*, $F(2,54) = 2.6$, $p = .08$, $= .03$, in that performance was best for compatible and worst for incompatible trials. Movement precision was also influenced by the compatibility between WM and movements goals, $F(2,54) = 7.4$, $p = .001$, $= .01$, as course adjustment errors were least frequent on compatible trials and most frequent on incompatible trials. This difference in the proportion of course adjustments also emerged as a benefit of compatible trials (relative to *equal priority*), $t(27) = 2.4$, $p = .022$, $d = 0.14$, $CI_{95\%}$ [0.02, 0.27], and a marginal cost of incompatible trials, $t(27) = 1.9$, $p = .069$, $d = 0.14$, $CI_{95\%}$ [0.00, 0.31] (**Fig. 2b**, right). When prioritized WM content was compatible with action goals, movement precision was improved, but when the two were incompatible, movement trajectories were more roundabout and required adjustment to arrive at the target.

Movement speeds. The compatibility of the cued WM item influenced *movement initiation*, $F(2,54) = 8.3$, $p < .001$, $= 0.008$, and *duration*, $F(2,54) = 5.0$, $p = .01$, $= .005$. There was a significant benefit of compatible cueing to *movement initiation*, $t(27) = 3.3$, $p = .003$, $d = 0.20$, $CI_{95\%}$ [0.06, 0.39], but no cost of incompatible cueing, $t(27) = 1.2$, $p = .25$, $d = 0.06$, $CI_{95\%}$ [-0.06, 0.17] (**Fig. 2c**, *left*). Conversely, there was a significant cost of incompatible cueing to *movement duration*, $t(27) = 2.4$, $p = .024$, $d = 0.10$, $CI_{95\%}$ [0.02, 0.21], but no benefit of compatible cueing, $t(27) = 1.2$, $p = 0.24$, $d = 0.07$, $CI_{95\%}$ [-0.04, 0.18] (**Fig. 2c**, *right*). When prioritized WM content matched the goals of a motor task, movement initiation was speeded, but when WM matched a motor task distractor, the movement itself took longer. This compatibility effect was reflected across the entire RT distribution (**SI Exp. 2 Results, Fig. S2**), rather than being driven by a small number of trials at the edges of the distribution.

Experiment 2 Discussion

While Experiment 1 manipulated the utility of WM content to the motor task, Experiment 2 used retrocues among two WM samples to manipulate their relative value to the WM test. Retrocues modulate neural signatures of WM representations (LaRocque et al., 2014), as well as pupil responses to WM content (Zokaei et al., 2019), and they are theorized to transform WM content into an "output-driving" state (Myers, Stokes, et al., 2017). We therefore predicted that retrocues would modulate the behavioral impact of WM, even on a task with a distinct goal from the WM task. As predicted, when multiple items were maintained in WM, the prioritized item influenced ongoing manual actions (and was also remembered better). Like Experiment 1, however, movement speeds suggested that prioritized WM content ignited movements toward

WM-compatible locations, but resulted in circuitous and slower movement paths on incompatible trials.

Even in a task context where WM goals have no predictive relationship to action goals (unlike Experiment 1), activated WM content can influence actions. Because Experiment 2 manipulated the relative priority status of two concurrently-maintained WM items, the findings suggest that the activation state of WM determines its sway over behavior. Theories of visual WM and attention have proposed that the activation status of a visual WM representation should determine whether it biases externally oriented visual attention (Olivers et al., 2011). The current findings suggest that this theoretical framework can be applied to understand the relationship between verbal WM and motor behavior as well.

However, retrocues were 90% valid in Experiment 2, which could have encouraged multiple strategies, like dropping the uncued item from memory entirely, or actively maintaining both items in case the uncued one were tested. Therefore, the distinction in attentional state between cued and uncued WM items was still ambiguous. In order to test competing theories about whether or not uncued (i.e., "unattended") WM content influences behavior, we devised an additional experiment in which uncued WM content always had to be maintained for later use. Experiment 3 will therefore provide a stronger manipulation of attentional prioritization within WM, so we can test the impact of attended versus "unattended" WM content and clarify the conditions that trigger WM biasing of action.

Experiment 3: Double retrocue manipulation

Methods

Rationale. This third (and final) experiment combines elements of Experiments 1 and 2 to examine the contributions of several distinct modes of WM modulation and influence over actions. Experiment 1 manipulated the task-level predictive relationship between WM and motor goals, while Experiment 2 manipulated the item-level likelihood that a given stimulus would need to be remembered. Experiment 3 combines these two manipulations, in a between-subjects design, to examine whether they evoke distinct or overlapping influences on behavior. This experiment again used a trial-by-trial retrocue procedure, but the task-level predictability between WM and motor goals was also manipulated between groups. Experiment 3 further employed two 100% valid retrocue phases (**Fig. 3a**). Thus, the cued item was known with certainty to be the one that should be prioritized in the first phase, but the uncued item still had to be retained because it was likely to become relevant again in the second phase. This allowed us to additionally test whether an unprioritized item imparts any trace on concurrent behavior when it may become relevant later.

Participants. Two new groups of participants were recruited using the same guidelines and IRB approval as Experiments 1 and 2. After 3 exclusions for below threshold accuracy (60%), Experiment 3 analyses included 30 (out of 30) participants in a "middle compatibility" group condition (12 male; mean age = 20.8 y, range = 18-30), and 29 (out of 32) participants in a "low compatibility" group condition (7 male; mean age = 20.9 y, range = 18-30).

Procedure. Like Experiment 2, two WM sample words were presented on each trial. After a delay of 1,500 ms, a retrocue (*1, 2,* or *X*) was displayed, but informative retrocues (*1* or *2)* were 100% valid indicators of which item would be probed, while an *X* indicated *equal priority*. After a second delay of 2,000 ms, participants completed the same motor task as in Experiments 1 and 2. Then the WM probe word appeared centrally until a response or a 3,000 ms deadline. On informative retrocue trials, participants compared the probe to the cued item. On *equal priority* trials, participants indicated whether the probe matched either item in WM. After an intermediate delay of 1,500 ms, the second trial phase began and participants were presented with another retrocue, motor task, and WM probe (**Fig. 3a**). Phase 2 followed the same structure and timing as Phase 1. After the second probe, a fixation cross appeared to indicate the start of a new trial. Participants completed one 8-trial practice block (with feedback), before completing 7 experimental blocks of 24 trials (without feedback).

The retrocue conditions (*1, 2,* or *X*) were evenly distributed, counterbalanced within each block and across trial phases, and each equally likely in the first and second phases. Like Experiment 2, the compatibility conditions are labeled depending on which of the two WM items was cued. However, the ratio of compatible to incompatible trials across the experiment was manipulated between groups. One experimental group was administered a version where, like Experiment 2, *compatible*, *incompatible*, and *equal priority* trials occurred equally often. Therefore 1/3 of all trials were compatible, and any given trial was equally likely to be one of the three conditions. Because half of all informatively cued trials were compatible, this is most akin to the middle (50%) compatibility condition from Experiment 1 and is therefore referred to as the "middle compatibility" group. The other group was administered a lower compatibility task version, wherein only one quarter of informatively cued trials were compatible, and therefore only 1/6 of all trials were compatible. This condition is referred to as the "low compatibility" group (**Fig. 3b**, *right*). Within a group, the proportion of compatible trials remained fixed across blocks. However, because incompatible trials could be categorized into subtypes that had an unequal distribution (described below), and the longer running time of the two-phase trial necessitated fewer trials per block, the incompatible subtypes could not be evenly divided for each block. Therefore, trial numbers of each condition were slightly variable across blocks and participants (+/-1 occurrence of each trial type in each block). In total, participants in the "middle compatibility" group completed 56 trials (+/- 7) per condition (*compatible* / *incompatible* / *equal priority* trials). Participants in the "low compatibility" group completed 28 compatible trials, 84 incompatible, and 56 equal priority trials.

In addition to assessing performance on the compatibility conditions that were tested in Experiment 2, here the experiment was designed to also test the behavioral impact of deprioritized or "unattended" WM content. On trials where the cued item was incompatible, the uncued item was selected from one of the three remaining words and could therefore be either compatible or incompatible with the movement direction. The label "fully incompatible with WM" will be used to describe trials where both the cued and uncued WM items were incompatible (2/3 incompatible trials), while "compatible with uncued WM" will be used to described trials where the cued item was incompatible but the uncued item was compatible (1/3 incompatible trials). The label "compatible with active WM" will describe trials where the cued item is compatible and the uncued item is necessarily incompatible, as it is impossible to have more than one compatible item in this task design. If deprioritized WM items have any impact on behavior, then "compatible with

uncued WM" trials—when an unprioritized WM item is compatible with the movement direction—should be faster than "fully incompatible with WM" trials.

Quality control criteria. As with Experiments 1 and 2, nonresponse trials for the WM probe, outliers for the motor response speed (> 3 s.d. from subject mean), and inaccurate motor task responses were excluded. Because Experiment 3 included two motor and WM phases, trimming criteria applied to both trial phases. In the "middle compatibility," group, 6.1% of total trials were excluded as response speed outliers, 0.3% as nonresponse trials, and 3.6% as response errors, while in the "low compatibility" group, 5.9% were excluded as response speed outliers, 0.5% as nonresponses, and 4.6% as response errors.

Analysis strategy: We first analyzed the compatibility conditions that were tested in Experiment 2, to assess whether the basic compatibility effects replicated in this modified task design. We conducted 3 (*trial compatibility*: compatible, equal priority, incompatible) × 2 (*task group*: middle vs. low compatibility) repeated measures ANOVAs on all Phase 1 performance measures. Trial type was modeled as a within-subjects factor and task compatibility as a between-subjects factor. We decomposed any significant interactions with follow-up one-way ANOVAs of trial compatibility, separately for each task group. Like Experiment 2, we further decomposed any main effects by examining benefits and costs (relative to *equal priority* trials). This task employed a second retrocue and motor phase in order to ensure that uncued WM content was still remembered for the later test. Although we had no hypotheses about performance in Phase 2, we also ran a full model of the experiment that included task phase. However, the task phase factor showed no main effects or interactions with either compatibility or task group (**SI Exp. 3 Results**). We also examined whether Phase 1 item priority levels impacted Phase 2 motor behavior, categorizing trials based on the interaction between Phase 1 and Phase 2 attentional states (**SI Exp. 3 Results, Fig. S4**). There were no significant effects of changing priority status, therefore the primary analyses below focus on Phase 1 alone.

To test the impact of deprioritized WM content, we calculated difference scores from the *equal priority* condition for each compatibility sub-condition. We conducted 3 (*trial compatibility*: compatible with active WM, compatible with uncued WM item, fully incompatible with WM) × 2 (*task group*: middle vs. low compatibility) repeated measures ANOVAs on these difference scores, for each movement speed measure. We also decomposed any significant interactions with follow-up one-way ANOVAs of trial compatibility, separately for each task group. One participant was excluded from this analysis because too few trials remained in the "compatible with uncued WM item" condition (the least frequent trial type) after removal of outliers and motor errors.

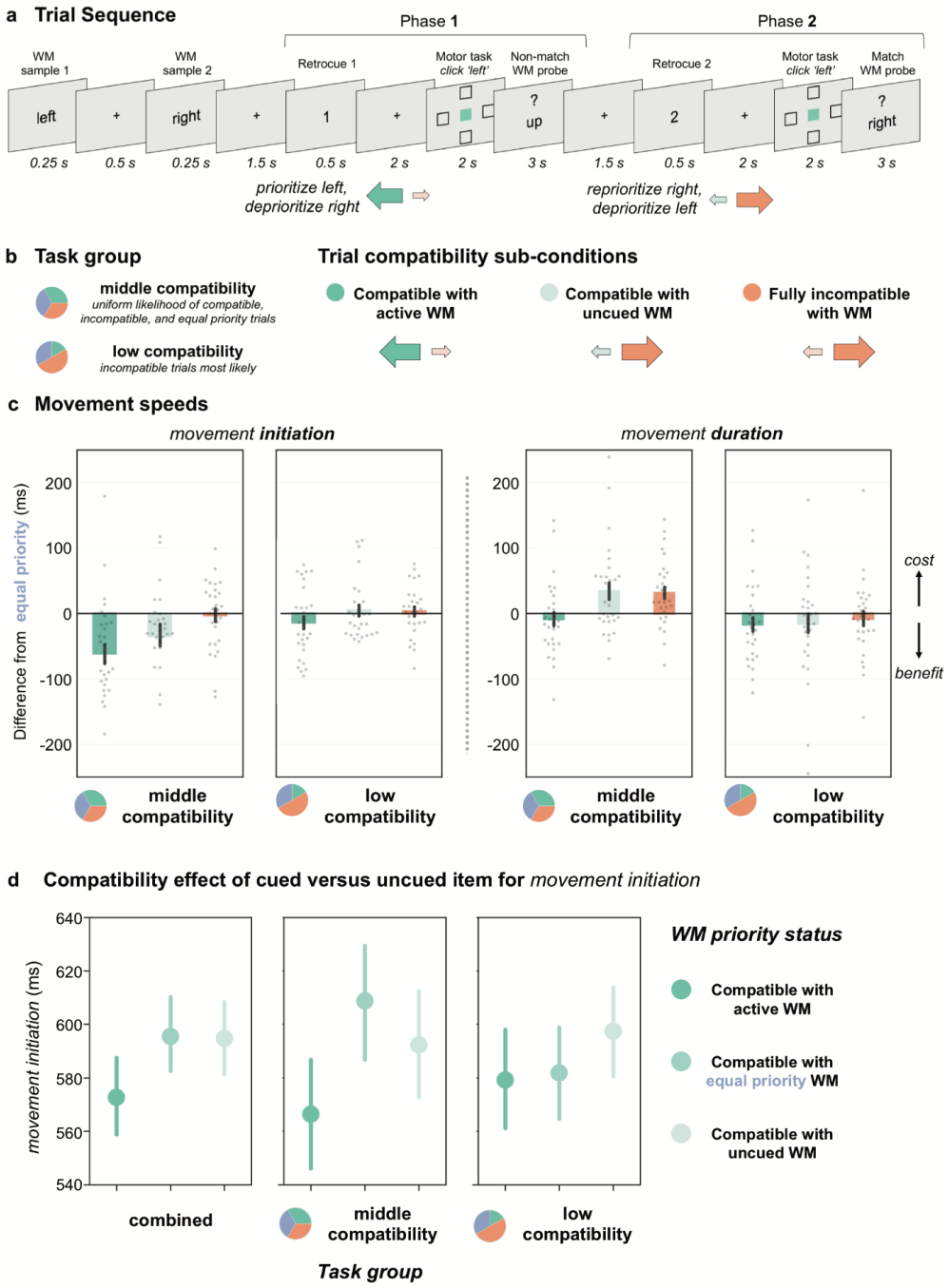


Fig. 3. Experiment 3 task design and results. **(a)** Example trial sequence. For each task phase, the cued WM sample could be either *compatible, uninformative (equal priority), or incompatible* with the movement direction. **(b)** Green and orange arrows illustrate the hypothetical priority levels of the WM samples in this example trial (left). Task predictability conditions differed across two task groups (right). (c) Cost and benefits relative to *equal priority* trials, split by 3 compatibility subtypes (*compatible with active WM / compatible with uncued WM / incompatible with WM*) for each movement measure (*initiation / duration*) and task phase (*Phase 1 / Phase 2*). **(d)** Phase 1 *movement initiation* times split by experiment task group (columns) and WM priority status (*compatible with active WM / compatible with equal priority WM / compatible with uncued WM*). Error bars represent SEM. Gray dots are data points from individual participants. In (c), one data point was excluded for visualization, but was included in all statistics.

Results

WM accuracy. WM probe accuracy was high in trial Phase 1 (92.4%), and showed an interaction between compatibility and task group, $F(2,114) = 3.24$, $p = .04$, $= .01$, but no main effects (compatibility: $F(2,114) = 1.05$, $p = .35$, $= .003$, task group: $F(1,57) = 0.08$, $p = .78$, $= .001$). This interaction was likely driven by a trending effect of compatibility in the low compatibility task group, $F(2,56) = 2.75$, $p = .07$, $= .01$, although there were no significant differences between any of the main trial types on Phase 1 WM accuracy. On trial Phase 2, WM probe accuracy was also high (93.1%), but with no main effect compatibility, $F(2,114) = 0.4$, $p = .65$, $= .001$, task group, $F(1,57) = 0.8$, $p = 0.4$, $= .01$, or interaction, $F(2,114) = 0.1$, $p = .49$ $= .0001$. Therefore, participants did retain both WM items for the second WM test.

Movement accuracy. Movement accuracy was high in trial Phase 1 (95.2%), but with no main effects of *cued compatibility*, $F(2,114) = 1.5$, $p = .23$, $= .004$, or task group, $F(1,57) = 0.6$, $p = .45$, $= .008$, nor an interaction, $F(2,114) = 1.6$, $p = .21$, $= .005$. Accuracy was slightly worse in Phase 2 (95.8%), but was unaffected by *cued compatibility*, $F(2,114) = 0.5$, $p = .59$, $= .002$, or task group, $F(2,114) = 1.9$, $p = .17$, $= .03$, and with no interaction, $F(2,114) = 2.2$, $p = .11$, $= .007$. Because of a programming error, cursor trajectory data are missing for Experiment 3, and the remaining analyses focus on movement speeds.

Movement speeds. There were no main effects of task group for either *movement initiation,* $F(1,57) = 0.2$, $p = .67$, $= .002$, or *duration,* $F(1,57) = 1.0$, $p = .31$, $= .02$, indicating that performance was comparable across groups. Replicating Experiment 2, however, there were strong effects of compatibility on both movement *initiation* and *duration,* which were reflected across the entire RT distribution for *initiation* (**SI Exp. 3 Results, Fig. S2**). Movement *initiation* displayed a main effect of compatibility, $F(2,114) = 13.6$, $p = < .001$, $= .02$, as well as an interaction with task group $F(2,114) = 4.4$, $p = .01$, $= .007$ (**Fig. S3)**. Follow-up one-way ANOVAs indicated that the compatibility effect was only present in the middle compatibility group, $F(2,58) = 13.0$, $p = < .001$, $= .04$, but not the low compatibility group, $F(2,58) = 2.1$, $p = .13$, $= .006$. In the middle compatibility group, there was a significant benefit of compatible cueing to *movement initiation*, $t(28) = 3.9$, $p < .001$, $d = 0.49$, $CI_{95\%}$ [0.23, 0.73], but no cost of incompatible cueing, $t(28) = 1.13$, $p = .27$, $d = 0.09$, $CI_{95\%}$ [-0.07, 0.24], much like Experiment 2.

Likewise, movement *duration* displayed a main effect of compatibility, $F(2,114) = 5.1$, $p = .008$, $= 0.007$, as well as an interaction with task group, $F(2,114) = 4.7$, $p = .01$, $= .007$. follow-up one-way ANOVAs indicated that the compatibility effect was only present in the middle compatibility group, $F(2,58) = 10.1$, $p < .001$, $= .02$, but not the low compatibility group, $F(2,58) = 1.17$, $p = .32$, $= .004$. Also like Experiment 2, there was no significant benefit of compatible cueing to *movement duration*, $t(28) = 0.9$, $p = .37$, $d = 0.09$, $CI_{95\%}$ [0.12, 0.31] but there was a cost of incompatible cueing, $t(28) = 3.6$, $p = .001$, $d = 0.28$, $CI_{95\%}$ [0.11, 0.51], in the middle compatibility group (**Fig. S3**). Therefore, for the middle compatibility group, the dissociable pattern of benefits and costs between movement *initiation* and *duration* replicated Experiment 2. However, these benefits and costs disappeared in the low compatibility task context, when WM content was unlikely to aid motor performance (1/6 of all trials).

These compatibility effects also persisted when the "total" compatibility of the WM sample set was held constant. That is, we repeated these ANOVAs including only trials where the WM sample contained at least one compatible item (i.e., *incompatible* trials where the uncued item was compatible, and *equal priority* trials where one of the 2 items was compatible). There were still main effects of compatibility on *movement initiation*, $F(2,114) = 4.0$, $p = .02$, $= .01$, and *movement duration*, $F(2,114) = 3.1$, $p = .05$, $= .007$ (**SI Exp. 3 Results**). Therefore, given trials that included both a compatible and an incompatible item, movement speeds were driven by which of the two were prioritized (**Fig. 3d**).

Additional analyses of each compatibility sub-condition provided further insight into the costs and benefits of WM content at different levels of attentional priority. We analyzed movement speed difference scores from *equal priority* trials as a function of whether (1) the cued WM item was compatible, (2) the uncued WM item was compatible, or (3) both WM items were incompatible (**Fig. 3b**). For *movement initiation*, there was a main effect of trial compatibility, $F(2,112) = 9.4$, $p < .001$, $= .06$, as well as an interaction between compatibility and task group, $F(2,112) = 3.0$, $p = .05$, $= .02$ (**Fig. 3c**, *left*). Follow-up tests showed that, in only the middle compatibility task group, there was a strong *initiation* benefit when the hand movement direction was compatible with the active (cued) WM item, $t(28) = 4.0$, $p < .001$, $d = 0.51$, $CI_{95\%}$ [0.24, 0.77]. There was also a marginal benefit compared to *equal priority* trials when the movement was compatible with the uncued WM item, $t(28) = 0.3$, $p = .051$, $d = 0.28$, $CI_{95\%}$ [0.05, 0.55], and no effect when both WM items were incompatible, $t(28) = 2.0$, $p = .78$, $d = 0.05$. Although there was a small benefit of a compatible uncued item (compared to equal priority), however, this condition did not differ from fully incompatible trials, $t(28) = 1.9$, $p = .07$, $d = 0.26$, $CI_{95\%}$ [0.03, 0.53]. Moreover, *movement initiation* was still significantly faster when the cued WM item was compatible, compared to the uncued item, in the middle compatibility group alone, $t(28) = 2.0$, $p = .05$, $d = 0.23$, $CI_{95\%}$ [0.05, 0.38], and combined across both task groups, $t(58) = 2.6$, $p = .01$, $d = 0.21$, $CI_{95\%}$ [-0.01, 0.48] (**Fig. 3d**).

To further probe a possible impact of relatively unattended WM content, we also tested whether *equal priority* trial performance varied according to the compatibility of the sample set. Indeed, *movement initiation* was faster on trials when the equal priority WM set included one compatible item versus two incompatible WM items, although neither was cued as most relevant – both for the middle compatibility group, $t(29) = 2.8$, $p = .009$, $d = 0.24$, $CI_{95\%}$ [0.06, 0.40], and marginally for the low compatibility group, $t(28) = 2.0$, $p = .06$, $d = 0.23$, $CI_{95\%}$ [-0.01, 0.50].

Finally, *movement duration* also displayed a main effect of compatibility when examining these trial sub-type difference scores, $F(2,112) = 5.3$, $p = .006$, $= .03$, as well as an interaction between compatibility and task group, $F(2,112) = 9.4$, $p < .001$, $= .06$ (**Fig. 3c**, *right*). As with the earlier experiments and analyses, in the middle compatibility group there was no benefit (relative to equal priority trials) when the movement direction was compatible with the active (cued) WM item, $t(28) = 0.8$, $p = .43$, $d = 0.08$, $CI_{95\%}$ [-0.03, 0.40]. However, there were costs when the WM content was fully incompatible with the movement direction, $t(28) = 3.3$, $p = .003$, $d = 0.29$, $CI_{95\%}$ [0.11, 0.56] as well as when the movement was compatible with the uncued WM item (but the cued item was still incompatible) $t(28) = 2.7$, $p = 0.01$, $d = 0.28$, $CI_{95\%}$ [0.09, 0.55]. In the case of movement duration, the unattended compatible content did not appear to impart any benefit. Again, these costs of incompatible cuing were eliminated in the low compatibility task group, when WM content was unlikely to aid motor performance on average.

Experiment 3 Discussion

Like Experiment 2, compatible WM content facilitated *movement initiation* while incompatible content slowed the motion itself (relative to equal priority trials). Here, we also show that deprioritized WM content can modestly influence action in certain contexts. *Movement initiation* was faster when there was a compatible item in the WM set (vs. two incompatible items), even if that item was not cued as relevant (either uncued or equal priority). Rather than an all-or-none influence of attentional prioritization, remembered items at lower priority may be maintained in a state that still influences behavior, but to a lesser extent than fully prioritized WM content. However, *movement duration* times were unaffected by deprioritized content, suggesting that accessory WM items may influence decision processes to start actions, but only prioritized content influences execution of the action itself.

Experiment 3 further supports the hypothesis that WM activation status modulates its impact on action. When two items were maintained, the attended one preferentially drove ongoing behavior. Item-level WM priority status also interacted with the task-level WM predictive utility to movement goals. When WM goals had a neutral relationship to motor goals (middle compatibility group, 1/3 compatible trials total), item-level activation status of WM content exerted a strong influence over behavior. However, in a task context when WM content was unlikely to help motor behavior (low compatibility group, 1/6 compatible trials total), both costs and benefits from WM content were eliminated. This suggests that the temporally-extended, higher-order task goals may take precedence and control the impact of phasic item-level attentional modulation.

General Discussion

Here, we examined how transient WM representations shape ongoing actions. We tested whether WM content influences movements executed during maintenance, and we probed the flexibility of that influence by manipulating the task-relevance of WM to either motor or mnemonic goals. Movements were less accurate, less precise, initiated later, and completed more slowly when they were incompatible with WM content. Across all three experiments, motor benefits of compatible WM content manifested in accelerated decision processes, while costs of incompatible content manifested in imprecise and time-consuming actions (**Table 1**). This effect on movement speeds depended on the priority level of WM representations, and was enhanced or diminished when WM content was predictably helpful or harmful to motor behavior, respectively. Visual WM content has long been shown to inadvertently influence perception and attention (Soto et al., 2008), and these findings now demonstrate that verbal WM exerts similar influences on motor behavior. However, rather than a monolithic effect on movement, WM content biases distinct stages of action execution in dissociable ways.

Theories of visual WM and attention have proposed that the 'activation state' of visual WM content will determine whether it incidentally biases perceptual and attentional processing (Olivers et al., 2011). A more recent proposal further distinguishes between neural and "functional" activation states, to suggest that the influence of WM will depend on its intended future use (Nobre & Stokes, 2019). The current results suggest that these frameworks can be applied to understand the incidental influence of verbal WM content on manual actions. These results support and augment theoretical conceptualizations of WM as intention or preparation for future actions (Fuster, 1990, 2004; Fuster & Alexander, 1971; Postle, 2006; Theeuwes et al., 2009). Here, verbal WM biases actions even when it is irrelevant (or detrimental) to the current task. This translation of WM content into (sometimes incorrect) actions confirms predictions of action control made by event coding theories, which propose a potentially automatic relationship between action selection and execution (Hommel, 2009). The present study also advances recent descriptions of a reciprocal influence between visual WM and planned motor behavior (van Ede, 2020), which can be observed in oculomotor (Hanning et al., 2016; Ohl & Rolfs, 2017; van Ede, Chekroud, & Nobre, 2019) and pointing movements (Heuer, Crawford, & Schubo, 2017). Here, we show that this framework may additionally extend to the influence between verbal WM and unplanned actions. Moreover, the task and trial-level modulations that we observed here also fit predictions from an intention-based reflexivity theory—that WM information in the focus of attention should exert the greatest influence over actions (Meiran et al., 2012). Therefore, several complementary theories of attentional and action control may be informed by the current findings.

While many studies have demonstrated the privileged status of attended WM content, the fate of unattended content remains debated (Myers, Chekroud, Stokes, & Nobre, 2017; Schneegans & Bays, 2017). Yet, unattended WM content evokes detectable neural traces in higher cortical regions (Christophel, Iamshchinina, Yan, Allefeld, & Haynes, 2018). Here, unprioritized (and presumably less active) WM content still modestly influenced movement initiation, suggesting that these representations may be sufficient to guide WM-based decisions. However, the results do show a graded influence of prioritization over WM content: uncued WM content influenced only movement initiation (not duration) and to a lesser extent than prioritized content. The observed distinction between movement initiation and duration also supports a theoretical

mechanism whereby prioritization reduces the time to access a given representation in WM, as suggested by drift diffusion modeling (Shepherdson, Oberauer, & Souza, 2017). That is, the graded effect of WM priority may be specific to movement initiation time because it influences decision processes involved in triggering a movement, rather than the mechanics that influence the duration of the movement itself.

The current data show that the WM-action bias can be modulated by modulating item representations, but superordinate task goals can also influence that item-level bias. Earlier theories of procedural WM suggest that executive systems mediate the interaction between immediate goals and a task set (Oberauer, 2009, 2010). This is consistent with the current finding that higher-order WM task context determines the degree to which WM influences motor behavior. WM item selection (or output gating) may work in tandem with task-level control functions that segregate competing demands in complex tasks (Badre & Nee, 2018). That is, WM goals may be prioritized over motor goals when WM is likely to aid motor performance, biasing motor decision-making toward WM (e.g., Experiment 1, high compatibility condition). Moreover, when WM and motor goals are equally weighted and may compete (e.g., a middle compatibility condition), item-level attentional modulation may tip the balance in favor of prioritized WM content. When WM is unlikely to aid motor behavior across the task, however, representations for WM and motor rules may be well-segregated (Verbruggen et al., 2018), minimizing leakage of WM content into action execution (e.g., Experiment 3, low compatibility condition). These findings may therefore extend the application of output gating and hierarchical control models to complex scenarios where WM content can trigger incorrect actions.

It remains unknown at which representational level these interactions occur, or whether the WM compatibility effect generalizes between content with less explicit overlap. In an adaptive system, it seems most likely that this interplay would in fact depend on the content domains being sufficiently related. Likewise, classic work on interference between WM and reading demonstrates that verbal and spatial recall are processed in a modality-specific manner, with the highest levels of interference when competing information is presented in the same domain (Brooks, 1968). The present results and related work therefore point to WM-motor interactions that may occur at an abstract, executive control level. For example, in a multi-component model of working memory, the spatially-focused visuospatial sketchpad and the verbally-focused phonological loop only have direct connections through a central executive system (Baddeley & Logie, 1999; Repovs & Baddeley, 2006). It may also be the case that verbal WM content for directions (e.g., ‘left’) generates neural signals similar to those for actions themselves, for instance, via priming (Mollo, Pulvermuller, & Hauk, 2016), or that the demand here to transform the symbolic motor cue into a direction resulted in an interfering verbal representation. Alternatively, participants here may have recoded verbal directions into visuo-spatial representations. However, the WM response in this task had no directional content, and used a different effector from the manual task, so we would not expect the WM response preparation to influence the motor task response. Yet, a visuo-spatial representation may still have been the most efficient or preferred maintenance strategy. So, having established some of the boundaries of the interplay between verbal WM and motor behavior, additional work should tease apart where those boundaries manifest representationally.

While it may seem intuitive that more task relevant WM content would exert the greatest impact on behavior, the compatibility of WM content to motor goals was typically unknown or

unlikely in most cases across these three Experiments. This was a simple WM task, performed with high accuracy, which should engage relatively modest attentional demands. Yet WM content imparted a dramatic influence on the course and speed of hand movements, producing a seemingly inadvertent impact over ongoing behavior. The modulation of this WM bias by task context, however, highlights the sensitivity of the system to temporally-extended regularities of the environment. Compatibility effects were consistently stronger when WM was most likely to help performance (on either the motor or WM task). Collectively, these results support the idea that occasional glitches in motor output stem from an adaptive WM system that adjusts to the correspondence between WM and other concurrent goals.

References

Baddeley, A. D., & Logie, R. H. (1999). Working Memory: The Multiple-Component Model *Models of Working Memory* (pp. 28-61).

Badre, D. (2012). Opening the gate to working memory. *Proc Natl Acad Sci U S A, 109*(49), 19878-19879. doi:10.1073/pnas.1216902109

Badre, D., & Nee, D. E. (2018). Frontal Cortex and the Hierarchical Control of Behavior. *Trends Cogn Sci, 22*(2), 170-188. doi:10.1016/j.tics.2017.11.005

Bays, P. M., & Taylor, R. (2018). A neural model of retrospective attention in visual working memory. *Cognitive Psychology, 100*, 43-52. doi:10.1016/j.cogpsych.2017.12.001

Belopolsky, A. V., & Theeuwes, J. (2011). Selection within visual memory representations activates the oculomotor system. *Neuropsychologia, 49*(6), 1605-1610. doi:10.1016/j.neuropsychologia.2010.12.045

Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J. C., Weissman, D. H., Notebaert, W., & Egner, T. (2019). Measuring Adaptive Control in Conflict Tasks. *Trends Cogn Sci*. doi:10.1016/j.tics.2019.07.002

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436.

Brooks, L. (1968). Spatial and Verbal Components of the Act of Recall. *Canadian Journal of Psychology, 22*(5), 349-368.

Carlisle, N. B., & Woodman, G. F. (2011). Automatic and strategic effects in the guidance of attention by working memory representations. *Acta Psychol (Amst), 137*(2), 217-225. doi:10.1016/j.actpsy.2010.06.012

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Curr Opin Behav Sci, 1*, 23-31. doi:10.1016/j.cobeha.2014.08.001

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., & Haynes, J. D. (2018). Cortical specialization for attended versus unattended working memory. *Nat Neurosci, 21*(4), 494-496. doi:10.1038/s41593-018-0094-4

Cools, R., Ivry, R. B., & D'Esposito, M. (2006). The Human Striatum is Necessary for Responding to Changes in Stimulus Relevance. *Journal of Cognitive Neuroscience, 18*(12).

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci, 18*, 193-222.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen, 141*(1), 2-18. doi:10.1037/a0024338

Fuster, J. (1990). Prefrontal Cortex and the Bridging of Temporal Gaps in the Perception-Action Cycle. *Ann N Y Acad Sci, 608*(1), 318-336.

Fuster, J. (2004). Upper processing stages of the perception-action cycle. *Trends Cogn Sci, 8*(4), 143-145. doi:10.1016/j.tics.2004.02.004

Fuster, J. (2015). Anatomy of the Prefrontal Cortex *The Prefrontal Cortex* (5 ed.): Elsevier.

Fuster, J., & Alexander, G. (1971). Neuron activity related to short-term memory. *Science, 173*, 652-654.

Hanning, N. M., Jonikaitis, D., Deubel, H., & Szinte, M. (2016). Oculomotor selection underlies feature retention in visual working memory. *J Neurophysiol, 115*(2), 1071-1076. doi:10.1152/jn.00927.2015

Heuer, A., Crawford, J. D., & Schubo, A. (2017). Action relevance induces an attentional weighting of representations in visual working memory. *Mem Cognit, 45*(3), 413-427. doi:10.3758/s13421-016-0670-3

Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychol Res, 73*(4), 512-526. doi:10.1007/s00426-009-0234-2

Ivry, R. B., & Spencer, R. M. (2004). The neural representation of time. *Curr Opin Neurobiol, 14*(2), 225-232. doi:10.1016/j.conb.2004.03.013

Kiyonaga, A., Egner, T., & Soto, D. (2012). Cognitive control over working memory biases of selection. *Psychon Bull Rev, 19*(4), 639-646. doi:10.3758/s13423-012-0253-7

LaRocque, J. J., Lewis-Peacock, J. A., & Postle, B. R. (2014). Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Human Neuroscience, 8*.

Lewis-Peacock, J. A., Drysdale, A., Oberauer, K., & Postle, B. R. (2011). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience, 24*(1), 61-79.

Mallett, R., & Lewis-Peacock, J. A. (2018). Behavioral decoding of working memory items inside and outside the focus of attention. *Ann N Y Acad Sci*. doi:10.1111/nyas.13647

Manohar, S. G., Zokaei, N., Fallon, S. J., Vogels, T. P., & Husain, M. (2019). Neural mechanisms of attending to items in working memory. *Neurosci Biobehav Rev, 101*, 1-12. doi:10.1016/j.neubiorev.2019.03.017

Meiran, N., Cole, M. W., & Braver, T. S. (2012). When planning results in loss of control: intention-based reflexivity and working-memory. *Front Hum Neurosci, 6*, 104. doi:10.3389/fnhum.2012.00104

Mollo, G., Pulvermuller, F., & Hauk, O. (2016). Movement priming of EEG/MEG brain responses for action-words characterizes the link between language and action. *Cortex, 74*, 262-276. doi:10.1016/j.cortex.2015.10.021

Myers, N. E., Chekroud, S. R., Stokes, M., & Nobre, A. C. (2017). Benefits of flexible prioritization in working memory can arise without costs. *Journal of Experimental Psychology*.

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends Cogn Sci*. doi:10.1016/j.tics.2017.03.010

Nobre, A. C., & Stokes, M. G. (2019). Premembering Experience: A Hierarchy of Time-Scales for Proactive Attention. *Neuron, 104*(1), 132-146. doi:10.1016/j.neuron.2019.08.030

O'Reilly, R., & Frank, M. J. (2005). Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Computation, 18*.

Oberauer, K. (2009). Design for a Working Memory *The Psychology of Learning and Motivation* (pp. 45-100).

Oberauer, K. (2010). Declarative and Procedural Working Memory: Common Principles, Common Capacity Limits? *Psychologica Belgica, 50*(3), 277-308.

Ohl, S., & Rolfs, M. (2017). Saccadic eye movements impose a natural bottleneck on visual short-term memory. *J Exp Psychol Learn Mem Cogn, 43*(5), 736-748. doi:10.1037/xlm0000338

Oliveira, F. T. P., & Ivry, R. B. (2008). The Representation of Action: Insights From Bimanual Coordination. *Current Directions in Psychological Science, 17*(2).

Olivers, C. N., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: when it guides attention and when it does not. *Trends Cogn Sci, 15*(7), 327-334. doi:10.1016/j.tics.2011.05.004

Park, Y. E., Sy, J. L., Hong, S. W., & Tong, F. (2017). Reprioritization of Features of Multidimensional Objects Stored in Visual Working Memory. *Psychological Science, 28*(12).

Postle, B. R. (2006). Working Memory as an Emergent Property of the Mind and Brain. *Neuroscience, 139*(1), 23-38.

Repovs, G., & Baddeley, A. (2006). The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience, 139*(1), 5-21. doi:10.1016/j.neuroscience.2005.12.061

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science, 354*(6316), 1136-1139. doi:10.1126/science.aah7011

Schneegans, S., & Bays, P. M. (2017). Neural Architecture for Feature Binding in Visual Working Memory. *J Neurosci, 37*(14), 3913-3925. doi:10.1523/JNEUROSCI.3493-16.2017

Schwartz, M. F. (1995). Re-examining the Role of Executive Functions in Routine Action Production. *Ann N Y Acad Sci, 769*.

Shepherdson, P., Oberauer, K., & Souza, A. S. (2017). Working Memory Load and the Retro-Cue Effect: A Diffusion Model Account. *J Exp Psychol Hum Percept Perform*. doi:10.1037/xhp0000448

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends Cogn Sci, 12*(9), 342-348. doi:10.1016/j.tics.2008.05.007

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Atten Percept Psychophys, 78*(7), 1839-1860. doi:10.3758/s13414-016-1108-5

Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring Latent Visual Working Memory Representations in Human Cortex. *Neuron, 91*(3), 694-707. doi:10.1016/j.neuron.2016.07.006

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci, 19*(7), 394-405. doi:10.1016/j.tics.2015.05.004

Theeuwes, J., Belopolsky, A., & Olivers, C. N. (2009). Interactions between working memory, attention and eye movements. *Acta Psychol (Amst), 132*(2), 106-114. doi:10.1016/j.actpsy.2009.01.005

van Ede, F. (2020). Visual working memory and action: Functional links and bi-directional influences. *Visual Cognition*, 1-13. doi:10.1080/13506285.2020.1759744

van Ede, F., Chekroud, S. R., & Nobre, A. C. (2019). Human gaze tracks attentional focusing in memorized visual space. *Nature Human Behaviour*. doi:10.1038/s41562-019-0549-y

van Ede, F., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2019). Concurrent visual and motor selection during visual working memory guided action. *Nat Neurosci, 22*(3), 477-483. doi:10.1038/s41593-018-0335-6

van Loon, A. M., Olmos-Solis, K., Fahrenfort, J. J., & Olivers, C. N. (2018). Current and future goals are represented in opposite patterns in object-selective cortex. *Elife, 7*. doi:10.7554/eLife.38677

van Moorselaar, D., Theeuwes, J., & Olivers, C. N. (2014). In competition for the attentional template: can multiple items within visual working memory guide attention? *J Exp Psychol Hum Percept Perform, 40*(4), 1450-1464. doi:10.1037/a0036229

Verbruggen, F., McLaren, R., Pereg, M., & Meiran, N. (2018). Structure and Implementation of Novel Task Rules: A Cross-Sectional Developmental Study. *Psychol Sci, 29*(7), 1113-1125.

Wallis, G., Stokes, M., Cousijn, H., Woolrich, M., & Nobre, A. C. (2015). Frontoparietal and Cingulo-opercular Networks Play Dissociable Roles in Control of Working Memory. *J Cogn Neurosci, 27*(10), 2019-2034. doi:10.1162/jocn_a_00838

Wolff, M. J., Ding, J., Myers, N. E., & Stokes, M. (2015). Revealing hidden states in visual working memory using electroencephalography. *Frontiers in Systems Neuroscience, 9*.

Wolff, M. J., Jochim, J., Akyurek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nat Neurosci*. doi:10.1038/nn.4546

Zokaei, N., Board, A. G., Manohar, S. G., & Nobre, A. C. (2019). Modulation of the pupillary response by the content of visual working memory. *Proc Natl Acad Sci U S A, 116*(45), 22802-22810. doi:10.1073/pnas.1909959116

Zokaei, N., Manohar, S., Husain, M., & Feredoes, E. (2014). Causal evidence for a privileged working memory state in early visual cortex. *J Neurosci, 34*(1), 158-162. doi:10.1523/JNEUROSCI.2899-13.2014

Chapter 4: Tertiary sulcal morphology links prefrontal anatomy and function

"... Our knowledge of corticocortical connectivity in the human, however, is still lagging behind. Given the ever more pressing evidence from brain imaging studies, which demonstrates the widely distributed and variable character of cortical networks, makes it imperative to further scrutinize corticocortical connectivity in the human in order to better understand the complexities of cognition."

- Fuster, *The Prefrontal Cortex* (Ch. 2)

This chapter contains previously published material from the following work, and permissions have been obtained from all co-authors for inclusion in this document. Supplemental information can also be found at the following reference:

Miller, J.A., Voorhies, W.I., Lurie, D.J., D'Esposito, M., Weiner, K.S. Overlooked tertiary sulci serve as a meso-scale link between microstructural and functional properties of human lateral prefrontal cortex. *Journal of Neuroscience* (2021) https://doi.org/10.1523/JNEUROSCI.2362-20.2021

Abstract

Understanding the relationship between neuroanatomy and function in portions of cortex that perform functions largely specific to humans such as lateral prefrontal cortex (LPFC) is of major interest in systems and cognitive neuroscience. When considering neuroanatomical-functional relationships in LPFC, shallow indentations in cortex known as tertiary sulci have been largely unexplored. Here, by implementing a multi-modal approach and manually defining 936 neuroanatomical structures in 72 hemispheres (in both males and females), we show that a subset of these overlooked tertiary sulci serve as a meso-scale link between microstructural (myelin content) and functional (network connectivity) properties of human LPFC in individual participants. For example, the posterior middle frontal sulcus (pmfs) is a tertiary sulcus with three components that differ in their myelin content, resting state connectivity profiles, and engagement across meta-analyses of 83 cognitive tasks. Further, generating microstructural profiles of myelin content across cortical depths for each pmfs component and the surrounding middle frontal gyrus (MFG) shows that both gyral and sulcal components of the MFG have greater myelin content in deeper compared to superficial layers and that the myelin content in superficial layers of the gyral components is greater than sulcal components. These findings support a classic, yet largely unconsidered theory that tertiary sulci may serve as landmarks in association cortices, as well as a modern cognitive neuroscience theory proposing a functional hierarchy in LPFC. As there is a growing need for computational tools that automatically define tertiary sulci throughout cortex, we share pmfs probabilistic sulcal maps with the field.

Significance statement

Lateral prefrontal cortex (LPFC) is critical for functions that are thought to be specific to humans compared to other mammals. However, relationships between fine-scale neuroanatomical structures largely specific to hominoid cortex and functional properties of LPFC remain elusive. Here, we show that these structures, which have been largely unexplored throughout history, surprisingly serve as markers for anatomical and functional organization in human LPFC. These findings have theoretical, methodological, developmental, and evolutionary implications for improved understanding of neuroanatomical-functional relationships not only in LPFC, but also in association cortices more broadly. Finally, these findings ignite new questions regarding how morphological features of these neglected neuroanatomical structures contribute to functions of association cortices that are critical for human-specific aspects of cognition.

Introduction

Understanding how anatomical structures of the brain support functional gradients and networks that perform computations for human-specific aspects of cognition is a major goal in systems and cognitive neuroscience. Of the many anatomical structures to target, lateral prefrontal cortex (LPFC) is expanded in the human brain relative to non-human primate species commonly used in neuroscience research, such as rhesus macaques (Semendeferi et al., 2002; Donahue et al., 2018; Barrett et al., 2020), and is particularly important given its central role in cognitive control and goal-directed behavior (Miller and Cohen, 2001; Szczepanski and Knight, 2014). Major progress has been made in understanding the relationship between the functional organization and the large-scale cortical anatomy of human LPFC. For example**,** previous findings support a hierarchical functional gradient organized along the rostral-caudal anatomical dimension of LPFC spanning several centimeters (Badre and D'Esposito, 2009; Nee and D'Esposito, 2016; Demirtas et al., 2019). Beyond this large-scale organization of human LPFC, it is largely unknown if more fine-grained structural-functional relationships exist. Thus, to begin to fill this gap in knowledge, we sought to answer the following question in the present study: Do individual differences in fine-grained morphological features of LPFC shed light on microstructural and functional properties of LPFC?

An important morphological feature of cortex is the patterning of the indentations, or sulci. Indeed, 60-70% of the cortex is buried in sulci and some sulci serve as landmarks that identify different cortical areas, especially in primary sensory cortices (Van Essen and Dierker, 2007; Zilles et al., 2013). In these cases, merely identifying a sulcus provides functional insight (Hinds et al., 2008). Despite this widely replicated relationship between sulcal morphology and functional representations in primary sensory cortices, much less is known regarding the predictability between shallow, tertiary sulci and functional representations in association cortex, especially LPFC. A classic theory proposed by Sanides (1964) hypothesized that the late emergence and protracted development of tertiary sulci may co-occur with microstructural and functional features of association cortices, along with cognitive functions such as sustained attention and "active thinking" (Sanides, 1964) that also develop fully after adolescence (Fisher, 2019).

However, at least two factors have prevented the examination of tertiary sulci relative to anatomical and functional organization in human LPFC. First, tertiary sulci are presently excluded from nearly all published neuroanatomical atlases because classic anatomists could not discriminate tertiary sulci from indentations produced by veins and arteries on the outer surface of the cerebrum in post-mortem tissue, which is considered the gold standard of anatomical research (Weiner et al., 2018). Consequently, tertiary sulci within the posterior middle frontal gyrus (MFG) were either undefined in classic atlases or conflated with more anterior structures (**Figure 1**; (Miller et al., 2020a)). Second, the majority of human functional magnetic resonance imaging (MRI) studies of LPFC implement group analyses on average brain templates. As shown in **Figure 1**, averaging cortical surfaces together causes tertiary sulci in LPFC to disappear, especially within the posterior MFG.

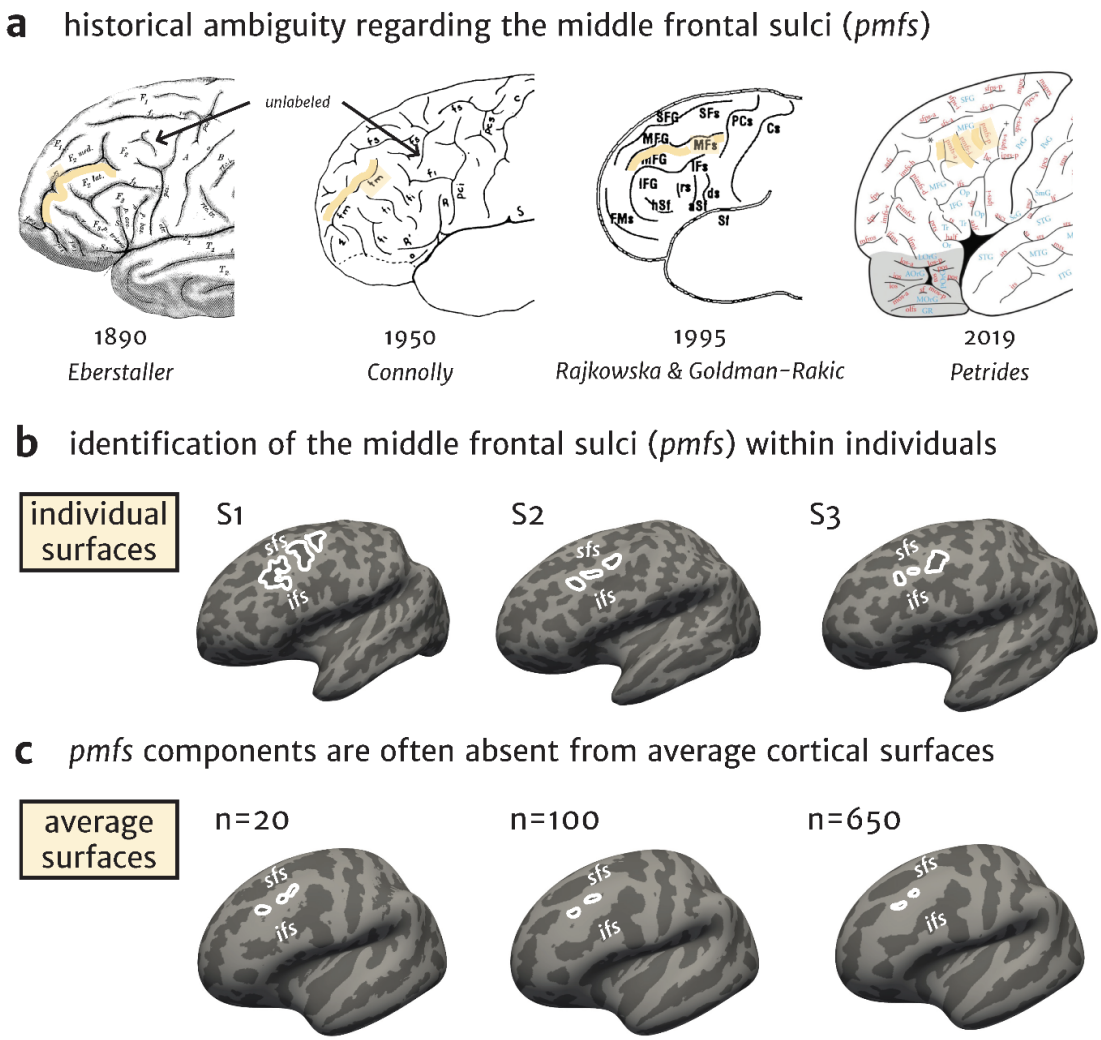


Figure 1. A synopsis of ambiguity regarding sulcal definitions in the human posterior middle frontal gyrus over the last 130 years. Classic and modern schematics of the sulcal patterning in human lateral prefrontal cortex (LPFC). (**a**) Sulci in the middle frontal gyrus are labeled in yellow on classic and modern schematics of human LPFC. Historically, anatomists had previously either (1) not labeled the sulci within the location of the modern *pmfs* (first two images; arrow indicates depicted, but unlabeled sulcal components) (Eberstaller, 1890; Connolly, 1950) or 2) included these sulci in the definition of the posterior portion of the frontomarginal sulcus (third image; (Rajkowska and Goldman-Rakic, 1995)). The most recent schematic (fourth image, **adapted from Petrides, 2019**) proposes that the *pmfs* is separate from the intermediate frontal sulcus (*imfs-h and imfs-v*, synonymous with the *frontomarginal sulcus*) and consists of three distinct components: posterior (*pmfs-p*), intermediate (*pmfs-i*), and anterior (*pmfs-a*). (**b**) Three individually labeled left hemispheres with the *pmfs* outlined in white. The *pmfs* is prominent within individual participants (**Extended Data Figure 2-1** for all participants). The superior and inferior frontal sulci (*sfs, ifs*) are labeled for reference above and below the middle frontal gyrus, respectively. (**c**) Average cortical surfaces show much smaller *pmfs* components compared to individual participants. As more participants are averaged together into templates, the *pmfs* disappears almost entirely, which is inconsistent with their prominence in individual hemispheres.

Here, we implemented a multi-modal approach demonstrating that identifying individual sulci in LPFC reveals that the *posterior middle frontal sulcus* (*pmfs*) serves as a meso-scale link between microstructural (myelin content) and functional (network connectivity) properties of human LPFC in individual participants. Specifically, after manually labeling LPFC tertiary sulci in 72 hemispheres based on a recently proposed labeling scheme (Petrides and Pandya, 2012; Petrides, 2019), we found that three components of the *pmfs* are dissociable based on myelin content, resting state functional connectivity profiles, and cognitive task activations. Moreover, the *pmfs* shows a distinct microstructural profile of myelin content across cortical depths from the surrounding MFG and distinct functional activations from the *intermediate frontal sulcus* (*imfs*). Together, these results not only provide important evidence that individual differences in LPFC sulcal patterning reflect meaningful differences in microstructural and functional properties, but also suggest that the *pmfs* serves as a bridge to Sanides' classic hypothesis.

Materials and Methods

In the sections below, we describe the data used and the analysis methods implemented in three separate sections: 1) the general approach and a description of the multi-modal datasets that were used, 2) a detailed description of the methodology used for sulcal labeling within individual participants, and 3) the calculation of anatomical and functional metrics.

General approach

We sought to characterize sulcal morphology at the individual level in the LPFC of the human brain. To implement this process, we manually defined sulci following the most recent and comprehensive proposed labeling of sulci in the frontal lobe (Petrides and Pandya, 2012; Petrides, 2019). As in our prior work (Weiner et al., 2014; Weiner et al., 2018), all sulci were defined in native space cortical surfaces and individual hemispheres, which enables the most accurate definition of tertiary sulci within *in vivo* MRI data.

Multi-modal HCP dataset

We analyzed a subset of the multi-modal MRI data available for individual participants from the Human Connectome Project (HCP). We began with the first 5 numerically listed HCP participants and then randomly selected 31 additional human participants from the HCP for a total of 36 individuals (17 female, 19 male, age range 22-36 years).

Anatomical T_1-weighted (T_1w) MRI scans (0.7 mm voxel resolution) were obtained in native space from the HCP database, along with outputs from the FreeSurfer pipeline slightly modified by the HCP (Dale et al., 1999; Fischl et al., 1999a; Glasser et al., 2013). Maps of the ratio of T_1-weighted and T_2-weighted scans, which is a measure of tissue contrast enhancement related to myelin content, were downloaded as part of the HCP 'Structural Extended' release. All additional anatomical metrics, which are detailed in the next section, were calculated on the full-resolution, native FreeSurfer (https://surfer.nmr.mgh.harvard.edu/) meshes (Dale et al., 1999; Fischl et al., 1999a; Fischl et al., 1999b).

Anatomical labeling and metrics

Manual sulcal labeling

Guided by a recent comprehensive proposal for labeling sulci in LPFC (Petrides, 2019), each sulcus was manually defined within each individual hemisphere on the FreeSurfer *inflated* mesh with *tksurfer*. The *curvature* metric in FreeSurfer distinguished the boundaries between sulcal and gyral components, and manual lines were drawn to separate sulcal components based upon the proposal by Petrides and colleagues (Amiez and Petrides, 2007; Petrides and Pandya, 2012; Petrides, 2019; Germann and Petrides, 2020), as well as the appearance of sulci across the *inflated*, *pial*, and *smoothwm* surfaces. We maintained the number of components for all tertiary sulci (e.g., the three components of the *posterior* m*iddle frontal sulcus - pmfs*) based on the proposal by Petrides and colleagues to test if each of these sulcal components could be defined in a relatively large sample size (N=72) of *in vivo* hemispheres. The labels were generated using a two-tiered procedure. The labels were first defined manually by J.M. and W.V. and then finalized by a neuroanatomist (K.S.W.). All anatomical labels for a given hemisphere were fully defined before any morphological or functional analysis of the sulcal labels was performed. The superior, inferior, posterior, and anterior boundaries of our cortical expanse of interest were the following sulci, respectively: (1) the anterior and posterior components of the *superior frontal sulcus*, (2) the *inferior frontal sulcus*, (3) the *central sulcus*, and (4) the horizontal (*imfs-h*) and vertical (*imfs-v*) intermediate frontal sulci. In each hemisphere, we first labeled the large primary sulci such as the central sulcus before labeling the secondary (e.g. *sfs*, *ifs*, *imfs*) sulci, and then we identified the tertiary sulcal components of the *pmfs*. Primary, secondary, and tertiary labels refer to the time in which the sulci emerge in gestation (Sanides, 1964; Chi et al., 1977; Welker, 1990; Armstrong et al., 1995). An example hemisphere with every sulcus labeled within these boundaries is shown in **Figure 2a**, and the *pmfs* sulcal components are plotted on each hemisphere in **Extended Data Figure 2-1**.

Quantification of sulcal depth and surface area

Sulcal depth was calculated from the native meshes generated by the FreeSurfer HCP pipeline. Raw values for sulcal depth (mm) were calculated from the sulcal fundus to the smoothed outer pial surface using a custom-modified version of a recently developed algorithm for robust morphological statistics building on the FreeSurfer pipeline (Madan, 2019). Surface area (mm^2) was generated for each sulcus through the *mris_anatomical_stats* function in FreeSurfer (Dale et al., 1999; Fischl et al., 1999a). We focused on sulcal depth as it is the main measurement that is used to discriminate tertiary sulci from primary and secondary sulci. Specifically, primary sulci are deepest, while tertiary sulci are shallowest, and secondary sulci are in between (Sanides, 1964; Chi et al., 1977; Welker, 1990; Armstrong et al., 1995). We also included surface area as tertiary sulci typically also have a reduced surface area compared to primary and secondary sulci.

Calculating T_1w/T_2w myelin index along an anterior-posterior dimension in LPFC

In order to test if there is a relationship between any of our sulci of interest and myelin content, we used an *in vivo* proxy of myelination: the T_1w/T_2w maps for each individual hemisphere (Glasser and Van Essen, 2011; Shams et al., 2019). To generate the T_1w/T_2w maps, two T1- and T2-weighted images from each participant were registered together and averaged as part of the HCP processing pipeline (Glasser et al., 2013). The averaging helps to reduce motion-related effects or blurring. Additionally (and as described in Glasser et al., 2013), the T_1w/T_2w images were bias-corrected for distortion effects with field maps. We averaged the T_1w/T_2w value across each vertex for each sulcus in order to test if the *pmfs* sulcal components are separable based on myelin content (**Figure 3**). We further sought to characterize the relationship between morphology and myelin by

determining if there was an anterior-posterior gradient of myelination across individual hemispheres. To do so, we first calculated the minimum geodesic distance of each vertex from the central sulcus. Geodesic distance was calculated on the *fiducial* surface using algorithms in the pycortex package (Gao et al., 2015). Then, we averaged across the vertices within each sulcus and tested for a linear relationship between average distance from the central sulcus and myelin content. To take advantage of each participant's individual data, we built a mixed linear model (random intercepts) in the *lme4* R package, using sulci and hemisphere as explanatory variables to correlate with average myelin content (**Figure 3**).

Sampling T_1w/T_2w myelin index across cortical depths

In order to investigate the microstructural profile of the *pmfs* across cortical layers, we generated nine surfaces from the outermost (*pial*) to the innermost (*white matter)* layers in all of the manually labeled hemispheres using an equivolumetric approach (Waehnert et al., 2014). We implemented the equivolume surface algorithm spanning nine cortical depths with the *surfacetools* Python package that builds on top of FreeSurfer (Dale et al., 1999) outputs: https://github.com/kwagstyl/surface_tools. The high-resolution T_1w/T_2w volumetric data in each HCP participant's native anatomical space were then sampled onto each equivolume surface using the FreeSurfer *mri_vol2surf* function to obtain a value of T_1w/T_2w at each cortical depth. The stability of depth profiles of T_1w/T_2w values extracted from individual regions was shown to be highest in the same HCP dataset when using a solution of 14 equivolume surfaces, with stability plateauing when using nine or more equivolume surfaces (Paquola et al., 2019). We compared the mean T_1w/T_2w value across depths for each participant in the manually defined *pmfs* components and the surrounding middle frontal gyrus (as defined by FreeSurfer parcellations (Destrieux et al., 2010), but with the *pmfs* components removed). We then conducted a repeated-measures ANOVA followed by post-hoc *t*-tests at each depth to test for differences in myelin content between the *pmfs* components and the *MFG* (**Figure 5**). Tests across each of the nine cortical depths were corrected for multiple comparisons at a familywise error (FWE) threshold of $p = 0.05/9$.

Cross-validation of sulcal location

In order to quantify the ability to predict the location of each sulcus across participants, we registered all sulcal labels to a common template surface (*fsaverage*) using cortex-based alignment (Fischl et al., 1999b). Similarity between each transformed individual label and the labels defined on *fsaverage* was calculated via the DICE coefficient, where X and Y are each label:

The cortex-based alignment algorithm aligns the surfaces based on sulcal depth and curvature metrics. We use the central sulcus as a proxy noise ceiling measurement for DICE coefficient values from other frontal sulci because it is a large and deep sulcus and is used in the surface registration algorithm that aligns cortical surfaces across participants (Fischl et al., 1999b).

Sulcal probability maps were calculated to describe the vertices with the highest alignment across participants for a given sulcus. A map was generated for each sulcus by calculating, at each vertex in the *fsaverage* hemisphere, the number of participants with that vertex labeled as the given sulcus, divided by the total number of participants. In order to avoid overlap among sulci, we then constrained the probability maps into *maximum probability maps* (MPMs) by only including vertices where (1) greater than 33% of participants included the given sulcal label and (2) the sulcus with the highest value of participant overlap was assigned to a given vertex. In a leave-one-participant out cross-validation procedure, we generated probability maps from n = 35 participants

and registered the probability map to the held-out participant's native cortical surface. This provided a measure of sulcal variability and prediction accuracy (**Figure 8**). This procedure also allows the identification of the *pmfs* sulcal components within held-out individual participants, reducing the extent of manual labeling necessary to identify this structure in future studies. Finally, the MPMs were used when analyzing meta-analytical functional data (described in the section *Cognitive Component Modeling*) and whole brain population receptive field data (**Figure 7**). The MPMs and code for alignment to new participants will be available on OSF with the publication of this paper.

Functional metrics

Resting-state network connectivity fingerprints

In order to test if the three *pmfs* sulcal components were functionally distinct from one another, we calculated and compared functional connectivity network fingerprints for each sulcus. Resting-state network parcellations for each individual participant were used from Kong et al. (2018), who generated individual network definitions by applying a hierarchical Bayesian network algorithm to produce maps for each of 17-networks (Yeo et al., 2011) in individual HCP participants. These data were calculated in the template HCP *fs_LR 32k* space. We resampled the network profiles for each participant onto the *fsaverage* cortical surface and, then, to each native surface using CBIG tools (https://github.com/ThomasYeoLab/CBIG). We then calculated the overlap of each *pmfs* sulcus in each participant with each of the 17 resting-state networks. We also separated the components of the *pmfs* and tested whether they showed similar or different network connectivity fingerprints using a 3-way repeated-measures ANOVA (sulcal component x network x hemisphere). Variability across individuals in the network profiles for each *pmfs* component was calculated by generating the Wasserstein metric (Earth Mover's Distance) between the resting-state network overlap values for each unique pair of participants (**Figure 5b**).

Cognitive component modeling

To further examine if the *pmfs-p, pmfs-i, and pmfs-a* are functionally distinct, we quantified the overlap between the *maximum probability maps* (MPMs) of each sulcal component and meta-analytic fMRI data from hundreds of experiments aligned to the *fsaverage* surface. Specifically, we quantitatively related the sulcal MPMs to vertex-wise maps for 14 cognitive components, which quantify how each vertex is recruited in a given set of cognitive operations across tasks and experiments (Yeo et al., 2015). We used a Bayesian method of expectation maximization to determine the combination of cognitive components that best fit each sulcal MPM. This resulted in a set of probabilities for each cognitive component for each sulcal map. We tested whether all sulci and the three components of the *pmfs* were distinguishable based upon these cognitive component loadings from a repeated-measures ANOVA (**Figure 6**).

Retinotopic response mapping

To determine if there was any correspondence between the manually labeled LPFC sulci and retinotopic representations, we analyzed a recent population receptive field mapping dataset (Benson et al., 2018). As these data were only available in a template (*fsaverage*) space, we used the predicted sulcal locations from probabilistic maps (as used in the cognitive components analysis) for these analyses (**Figure 7**). For each sulcus, we extracted the mean R^2 value (the percentage of variance in each vertex explained by the population receptive field model) across

participants for vertices that showed meaningful retinotopic responses (thresholded at $R^2 > 10\%$, as in (Mackey et al., 2017)).

Statistical methods

All repeated measures ANOVAs (including sphericity correction) and post-hoc t-tests were performed with the *afex* and *emmeans* R packages, imported into Python via *rpy2*. For each repeated measures ANOVA, cortical hemisphere and sulcus were used as within-subject factors. Effect sizes for each main effect and interaction were calculated and reported with the *generalized eta-squared* metric (Fritz et al., 2012). Mixed linear models were implemented in the *lme4* R package. Cortical surface files were loaded in and operated on in Python using the nilearn software: https://nilearn.github.io

Results

Before conducting our multimodal examination relating morphological features of tertiary sulci to microstructural and functional properties of LPFC, we first had to confront the contradictory nature of historic and modern definitions of sulci within the middle frontal gyrus (MFG). For example, sulcal definitions within the MFG vary in a) their nomenclature, b) the number of sulcal components depicted or acknowledged in schematics, c) the omission or inclusion of sulci within the posterior MFG, and d) the actual empirical data that is included to support the illustration of the sulcal patterning (**Figure 1**). To ameliorate these concerns and to either empirically support or to refute the generality of sulcal definitions within the posterior MFG, we apply a classic, multimodal approach that has been used to distinguish cortical areas from one another in order to determine sulcal definitions in the posterior MFG. Specifically, after identifying each sulcus within the posterior MFG based on recent proposals (Petrides and Pandya, 2012; Petrides, 2019), we use both anatomical and fMRI data to either support or refute the identification of individual sulci within this cortical expanse. Implementing this two-pronged approach, we first examined if the three components of the posterior middle frontal sulcus (*pmfs*) are consistently identifiable within individual hemispheres. And if so, we then tested if the three *pmfs* components are anatomically and functionally homogenous, or serve to identify anatomical and functional heterogeneity in LPFC. This approach supports the latter in which there are three anatomically and functionally distinct sulci within the posterior MFG: the posterior (*pmfs-p*), intermediate (*pmfs-i*), and anterior (*pmfs-a*) posterior middle frontal sulci.

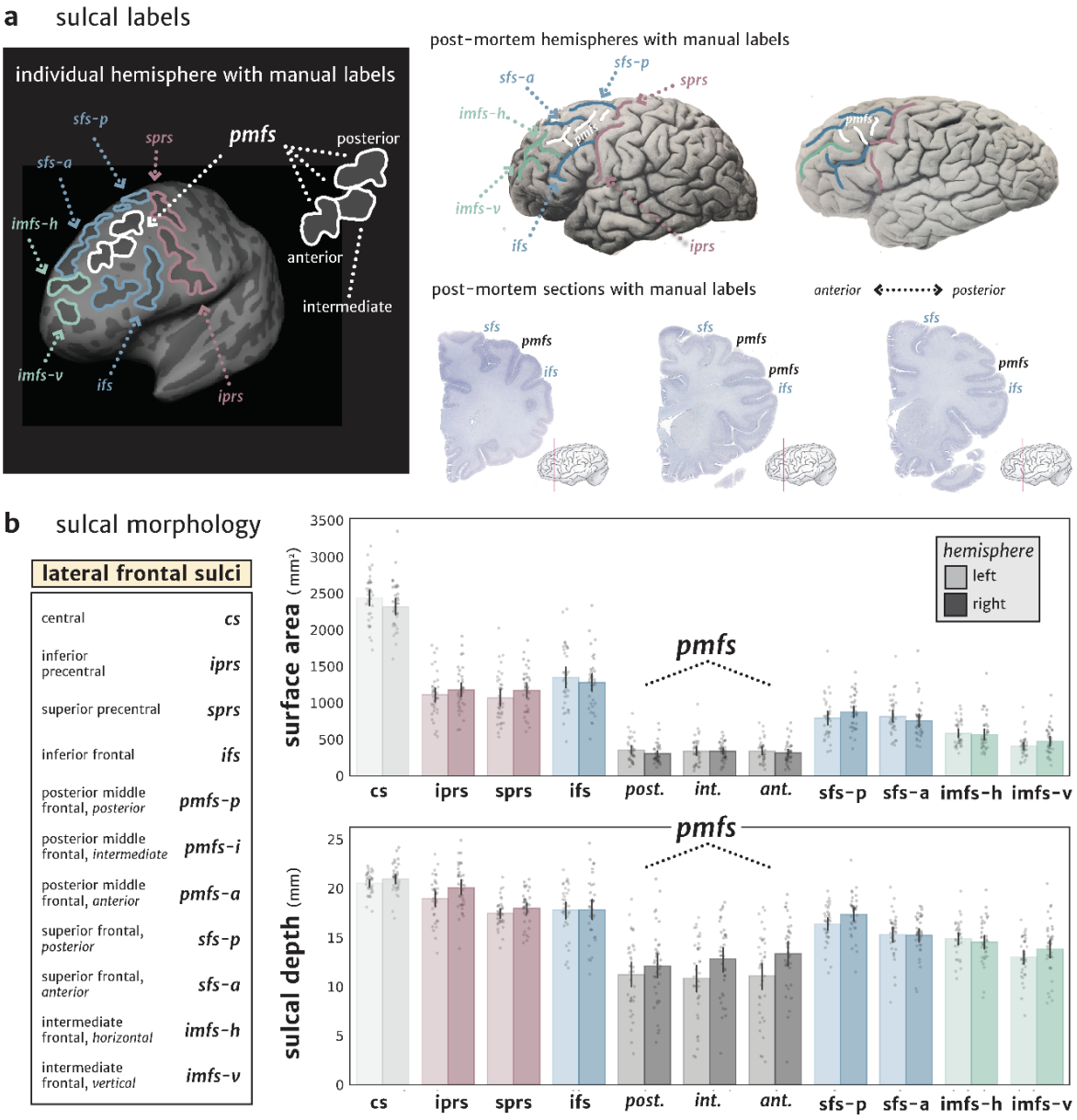


Figure 2. LPFC tertiary sulci are easily identifiable and characteristically shallow. (**a**) *Left:* an example inflated cortical surface of an individual left hemisphere in which the sulci examined in the present study are outlined and labeled (**Extended Data Figure 2-1** for all participants). Sulci are dark gray, while gyri are light gray. *Right:* Two post-mortem hemispheres (Retzius, 1896) and three histological sections (note that the pmfs components are referred to as "intermediate frontal sulcus" in the Allen Human Brain Atlas: https://atlas.brain-map.org/; (Ding et al., 2016)) showing that the *pmfs* sulci are also identifiable in post-mortem tissue samples. (**b**) Top: Surface area for each sulcus (ordered posterior to anterior) is plotted for each individual participant (gray circles), as well as the mean (colored bars) and 95% confidence interval (black line). Acronyms used for each LPFC sulcus are also included. Darker shades indicate right hemisphere values, while lighter shades indicate left hemisphere values. The three *pmfs* sulci have the smallest surface area of all LPFC sulci measured in the present study. Bottom: Same layout as above, but for sulcal depth (mm). The three *pmfs* sulci are the shallowest of the LPFC sulci measured here.

Three posterior middle frontal sulci (pmfs) are identifiable within individuals and are characteristically shallow

Before examining the sulcal patterning within the posterior MFG, we first identified reliable sulci (**Materials and Methods**: *manual sulcal labeling*) surrounding the MFG in both *in vivo* cortical surface reconstructions of MRI data and post-mortem brains (**Figure 2a**). Posteriorly, we identified the central sulcus (*cs*), as well as the superior (*sprs*) and inferior (*iprs*) pre-central sulci. Superiorly, we identified the anterior (*sfs-a*) and posterior (*sfs-p*) superior frontal sulci. Inferiorly, we identified the inferior frontal sulcus (*ifs*). Anteriorly, we identified the horizontal (*imfs-h*) and vertical (*imfs-v*) intermediate frontal sulci. The latter two sulci are consistent with Eberstaller's classic definition of the middle frontal sulcus, but have since been renamed (**Figure 1**; (Miller et al., 2020a)). Within the posterior MFG, we identified three sulci in every hemisphere (N=72). From posterior to anterior, the first sulcus (*pmfs-p*) is positioned immediately anterior to the *sprs* (**Figure 2a**, **Extended Data Figure 2-1**), and most commonly does not intersect other sulci (see **Table 1** for a summary of the morphological patterns, or types). The second sulcus (*pmfs-i*) is located immediately anterior to the *pmfs-p*, and typically aligns with the separation between the *sfs-a* and *sfs-p* components. The *pmfs-i* is most often independent (especially in the right hemisphere) or intersects (especially in the left hemisphere) the *pmfs-a*. Finally, the third sulcus (*pmfs-a*) is immediately anterior to the *pmfs-i*, inferior to the *sfs-a*, and posterior to the *imfs-h*. The *pmfs-a* most commonly intersects other sulci in the right hemisphere.

Table 1. Most common intersections of the pmfs components (morphological types)

Most common intersections	1st	2nd	3rd
pmfs-p	Independent	pmfs-i	iprs
lh	44.4%	22.2%	16.7%
	Independent	sfs-a	pmfs-i
rh	30.6%	30.6%	16.7%
pmfs-i	Independent	pmfs-p	pmfs-a
lh	47.2%	22.2%	16.7%
	pmfs-a	Independent	pmfs-p
rh	58.3%	27.8%	19.4%
pmfs-a	Independent	imfs-h	pmfs-i
lh	47.2%	38.9%	16.7%
	imfs-h	pmfs-i	Independent
rh	52.8%	50.0%	13.9%

Each sulcus is also identifiable within individual *in vivo* volumetric slices (Petrides, 2019) and in postmortem brains (**Figure 2**), which indicates that the computational process used to generate the cortical surface reconstruction in the MRI data does not artificially create these sulci within the MFG. Our results show that the *pmfs* is distinguishable from the *imfs*, which is in correspondence with the recent atlas from Petrides (2019), whereas the *pmfs* and *imfs* were often combined in classic sulcal atlases (Ono et al., 1990).

The two most identifying morphological features of the three *pmfs* sulci are their surface area and depth (**Figure 2b**). Each *pmfs* sulcus is of roughly equal surface area (**Figure 2b**, **Table 2**), which is smaller than the surface area of the other examined sulci in LPFC (**Figure 2b**, **Table 2**). A two-way repeated-measures ANOVA with factors sulcus and hemisphere yielded a main effect of sulcus ($F(5.78, 202.15) = 384.1$, $p < 0.001$, $= 0.84$) and no main effect of hemisphere ($F(1, 35) = 0.1$, $p = 0.77$). The depth of the three *pmfs* sulci are also the shallowest of the lateral PFC sulci examined (**Figure 2b, Table 1**). A two-way repeated-measures ANOVA with sulcus and hemisphere as factors yielded a main effect of sulcus ($F(3.15, 103.84) = 77.7$, $p < 0.001$, $= 0.55$), and a main effect of hemisphere ($F(1, 33) = 20.4$, $p < 0.001$, $= 0.02$) in which sulci were deeper in the right compared to the left hemisphere (**Figure 2b**, **Table 2**). Post-hoc tests show that, across hemispheres, the *pmfs-p* is shallower than all other sulci (p-values < 0.001, Tukey's adjustment), and the *pmfs-i* and *pmfs-a* are shallower than all other sulci except for the *imfs-v*. Taken together, three *pmfs* sulci are identifiable in individual hemispheres (**Figure 2**, **Extended Data Figure 2-1**) and distinguish themselves from other LPFC sulci based on their surface area and shallowness.

Table 2. Surface area and depth of the three pmfs components

	Surface area (mm^2)	Depth (mm)
pmfs-a		
lh	341.9 ± 154.8	11.1 ± 4.4
rh	315.4 ± 149.7	13.4 ± 3.7
pmfs-i		
lh	339.3 ± 191.7	10.9 ± 4.2
rh	337.8 ± 124.2	12.8 ± 3.8
pmfs-p		
h	353.6 ± 164.1	11.2 ± 3.8
rh	301.7 ± 133.2	12.1 ± 3.9

Values are mean ± SD.

The pmfs-p, pmfs-i, and pmfs-a are anatomically dissociable and reflect a larger rostro-caudal myelination gradient in LPFC

While the *pmfs-p*, *pmfs-i*, and *pmfs-a* are morphologically distinct from surrounding sulci (**Figure 2**), it is presently unknown if they are anatomically and functionally similar or distinct from one another. To test this, we first extracted and compared average MRI T_1w/T_2w ratio values from each sulcus. The T_1w/T_2w ratio is a tissue contrast enhancement index that is correlated with myelin content (**Figure 3a**; (Glasser and Van Essen, 2011; Shams et al., 2019)). We chose this index because myeloarchitecture is a classic criterion used to separate cortical areas from one another (Vogt and Vogt, 1919; Flechsig, 1920; Hopf, 1956; Dick et al., 2012). A two-way repeated-measures ANOVA with sulcus and hemisphere as factors yielded a main effect of sulcus ($F(1.76, 61.7) = 85.0$, $p < 0.001$, $= 0.39$) and a main effect of hemisphere ($F(1, 35) = 10.5$, $p = 0.003$, $= 0.05$) on myelin content, but no sulcus x hemisphere interaction ($F(1.73, 60.5) = 2.5$, $p = 0.10$). The differences in myelin across sulci were driven by the finding that T_1w/T_2w decreased from posterior to anterior across hemispheres: *pmfs-p* vs. *pmfs-i,* $t(70) = 9.75$, $p < 0.001$ (Tukey's post-hoc), *pmfs-i* vs. *pmfs-a*, $t(70) = 2.62$, $p = 0.029$, and *pmfs-p* vs. *pmfs-a*, $t(70) = 12.37$, $p < 0.001$. The right hemisphere also had higher myelin content overall in the *pmfs*, $t(35) = 3.25$, $p = 0.003$. Accordingly, the three sulcal components are differentiable based on myelin content in both hemispheres (**Figure 3b**).

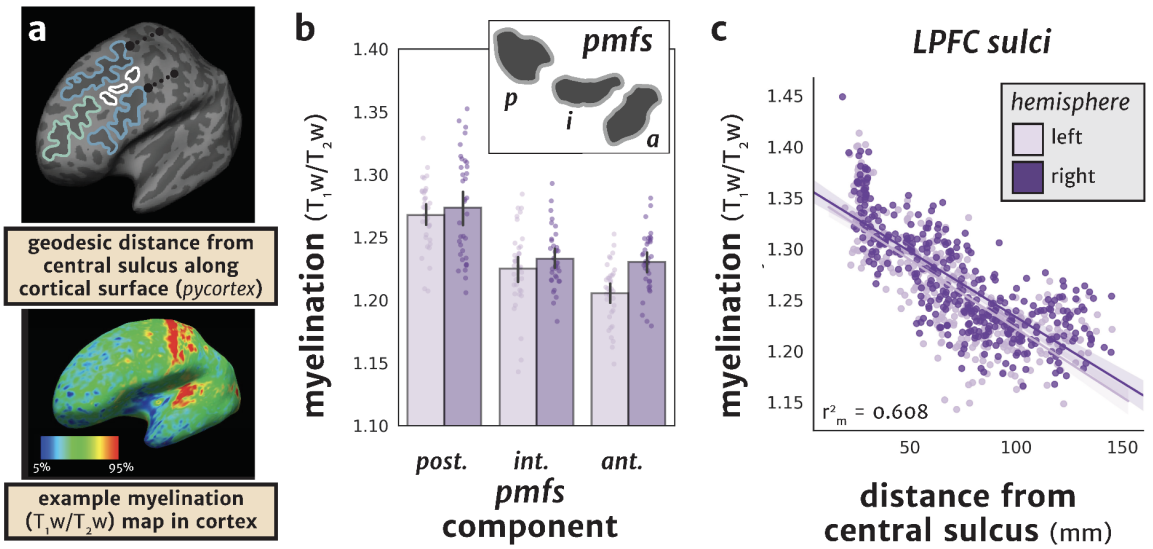


Figure 3. The *pmfs* sulci are anatomically differentiable based on myelin content. (**a**) Top: Schematic of the calculation of geodesic distance along the cortical surface. For each sulcus, the average distance of each vertex from the central sulcus was calculated (dotted black line; **Materials and Methods**). Bottom: an example T_1w/T_2w map in an individual participant in which 5-95% percentile of values are depicted. (**b**) T_1w/T_2w values (a proxy for myelin content) are plotted for each component of the *pmfs* for each individual participant (N = 36). Bars represent mean ± 95% CI, while each participant is depicted as a circle. Darker shades indicate right hemisphere values, while lighter shades indicate left hemisphere values. The components of the *pmfs* are differentiable based on myelin content, with a decrease from posterior to anterior across both hemispheres. (**c**) Scatterplot showing the negative relationship between distance from the central sulcus and the mean myelination value for all labeled sulci from each individual (N = 36 participants). The mixed linear model (**Materials and Methods**) with predictors of distance and hemisphere shows a marginal r^2 of 60.8%. Scatterplot is bootstrapped at 68% CI for visualization. (**d**) Scatterplot showing the mean T_1w/T_2w value for each sulcus as a function of distance (mm) from the central sulcus. Error bars for both the x- and y-axes represent S.E.M. (68% CI) across individuals (N = 36 participants). *Dark purple*: right hemisphere; *Light purple*: left hemisphere.

The rostro-caudal gradient among the *pmfs-p*, *pmfs-i*, and *pmfs-a* sulci is embedded within a larger rostro-caudal myelination gradient in lateral PFC. Specifically, modeling T_1w/T_2w content across frontal sulci as a function of distance from the central sulcus (**Figure 3c**) using a mixed linear model revealed a significant, negative effect of distance from the central sulcus along the rostral-caudal axis (= -0.001, z = -33.8, $p < 0.001$), with no differences between hemispheres (= -0.003, z = -0.8, $p = 0.4$). Together, our quantifications show that the *pmfs-p*, *pmfs-i*, and *pmfs-a* are embedded within a larger anatomical and functional hierarchical gradient in LPFC (see **Discussion** for further details).

The pmfs components show a microstructural profile across cortical layers that is distinct from the middle frontal gyrus (MFG)

Classic and modern findings show that there is generally more intracortical myelin in deeper cortical layers and that the depths of sulci often have less myelinated fibers than gyral crowns (Braitenberg, 1962; Sanides, 1972; Welker, 1990; Annese et al., 2004; Rowley et al., 2015). Building on this work, we sought to calculate microstructural profiles for myelin content across cortical depths for each *pmfs* component, as well as the gyral components of the MFG that surround them (**Figure 4**; **Materials and Methods**). To do so, we implemented equivolume algorithms to construct cortical surfaces within the gray matter. The depth profiles from equivolume surfaces have been used to investigate cortical laminar organization *in vivo* and correspond with those obtained from both *ex vivo* MRI data and post-mortem histological sections (Waehnert et al., 2014; Paquola et al., 2019).

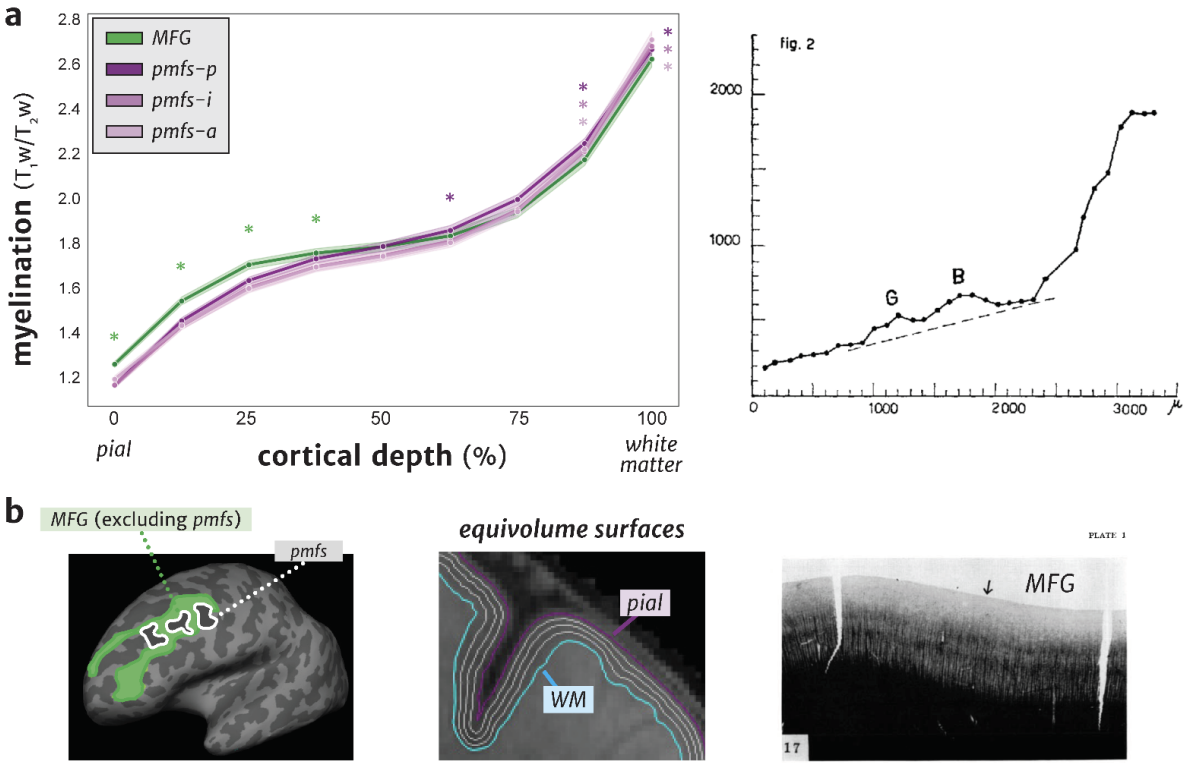


Figure 4. The *pmfs* sulci and middle frontal gyrus have differentiable myelin profiles across cortical depths. (**a**) Left: Tissue contrast enhancement (T_1w/T_2w metric, a proxy for myelin) at nine cortical depths, sampled from the outer gray matter (*pial*) to the gray/white matter boundary (*white matter*) using equivolume surfaces (**Materials and Methods**). The MFG (excluding the *pmfs*) has higher myelin content than all *pmfs* components in the upper cortical layers, while the *pmfs* components have higher myelin content in deeper layers. Shaded area represents bootstrapped 68% CI across participants. Green asterisks show significant statistical differences between the MFG and all *pmfs* components (MFG > *pmfs*), while purple asterisks show the reverse (*pmfs* > MFG; all tests FWE-corrected at $p < 0.05/9$). Right: Myelinated fiber density (y-axis) profile across cortical depths (x-axis) in post-mortem histological sections of the MFG, adapted from Braitenberg (1962). B: stria of Baillarger. G: stria of Gennari. Similar to our measurements, myelination increases from outer to inner layers within the MFG. (**b**) Left: Individual left hemisphere with the manually defined *pmfs* components (white) and the surrounding MFG (green) as defined by FreeSurfer

(Destrieux et al., 2010). We excluded the *pmfs* components from the MFG to test for anatomically distinct profiles. Middle: Example equivolume surfaces at five different cortical depths, from the *pial* to *white matter* surfaces, which were used to sample the T_1w/T_2w metric across depths. Right: Myelination stain of a post-mortem histological section of the MFG from Braitenberg (1962). Arrow: Location from which the myelinated fiber density profile in (**a**, right) was calculated.

The MFG and *pmfs* components show distinct microstructural profiles of myelin content across cortical depths. A three-way repeated-measures ANOVA with factors of structure (*pmfs-p, pmfs-i, pmfs-a, MFG*), cortical depth (*0%, 12.5%, 25%, 37.5%, 50%, 62.5%, 75%, 87.5%, 100%*), and hemisphere (*left, right*), yields main effects of structure ($F(2.26, 78.94) = 15.6$, $p < 0.001$, $= 0.007$), depth ($F(1.39, 48.49) = 1849.6$, $p < 0.001$, $= 0.84$), and a structure x depth interaction ($F(6.78, 237.43) = 78.5$, $p < 0.001$, $= 0.02$). This interaction between structure and depth did not differ by hemisphere ($F(4.69, 164.26) = 1.13$, $p = 0.35$, $= 0.02$), so subsequent analyses are collapsed across hemispheres. To determine which differences drive the distinct profiles in myelin content across cortical layers between the *pmfs* and MFG, we conducted post-hoc tests at each cortical depth (**Figure 4a**). The MFG had higher myelin content in each of the upper cortical depths (0%, 12.5%, 25%, 37.5%) compared to all of the *pmfs* components (all *p-values* < 0.001, FWE-corrected at $= 0.05/9$ for the 9 cortical depths). In the middle-to-deep layers (50%, 62.5%), the *pmfs-p* had higher myelin content than either the *pmfs-i* (50%: $t(105) = 6.4$, $p < 0.001$; 62.5%: $t(105) = 7.0$, $p < 0.001$) or *pmfs-a* (50%: $t(105) = 7.1$, $p < 0.001$; 62.5%: $t(105) = 8.1$, $p < 0.001$), and was even higher than the MFG (50%: $t(105) = 0.27$, $p = 0.99$; 62.5%: $t(105) = 3.7$, $p = 0.002$). At the deepest cortical layers, closest to the gray/white matter boundary, all three *pmfs* components showed increased myelin relative to the MFG. Specifically, the *pmfs-a* showed the highest myelin content in the deepest layers, but all three *pmfs* components displayed higher myelin than the MFG (all *p-values* < 0.001, FWE-corrected at $= 0.05/9$ for the 9 cortical depths). The profile of myelin content across cortical depths in the *pmfs* and MFG is also robust when comparing myelin content at a coarser (3 instead of 9) level of upper, middle, and lower depths (mean of depths within each bin): structure x depth interaction ($F(3.87, 135.4) = 127.4$, $p < 0.001$, $= 0.02$). Altogether, the *pmfs* differed from the MFG in microstructure across cortical layers, with lower myelin content in upper layers and higher myelin content in deeper layers. This surface-based sampling of cortical depths provides *in vivo* neuroimaging evidence for a microanatomical distinction of the *pmfs* from the surrounding MFG. Further, the depth profiles of T_1w/T_2w values within the MFG are similar to classic myeloarchitectural quantifications of the MFG (**Figure 4**).

The pmfs-p, pmfs-i, and pmfs-a exhibit different characteristic patterns of whole brain functional connectivity

To determine if the *pmfs-p, pmfs-i,* and *pmfs*-a are functionally distinct, we leveraged detailed individual functional parcellations of the entire cerebral cortex based on functional connectivity from a recently published study (Kong et al., 2018; **Figure 5a**). Importantly, this parcellation was conducted blind to both cortical folding and our sulcal definitions. Within each hemisphere in the same participants in which we generated manual sulcal labels, we generated a functional connectivity network profile (which we refer to as a "connectivity fingerprint"). For each sulcal component, we calculated the overlap between 17 functional networks (on the native hemisphere, based on the DICE coefficient; **Materials and Methods**). This technique generated a cortical topography reflective of the whole-brain connectivity patterns for each sulcal component (**Figure 5a, bottom**), and can be interpreted similarly to other studies of functional network variations (Gordon et al., 2017; Seitzman et al., 2019), as a trait-like connectivity profile for each *pmfs* component within each participant.

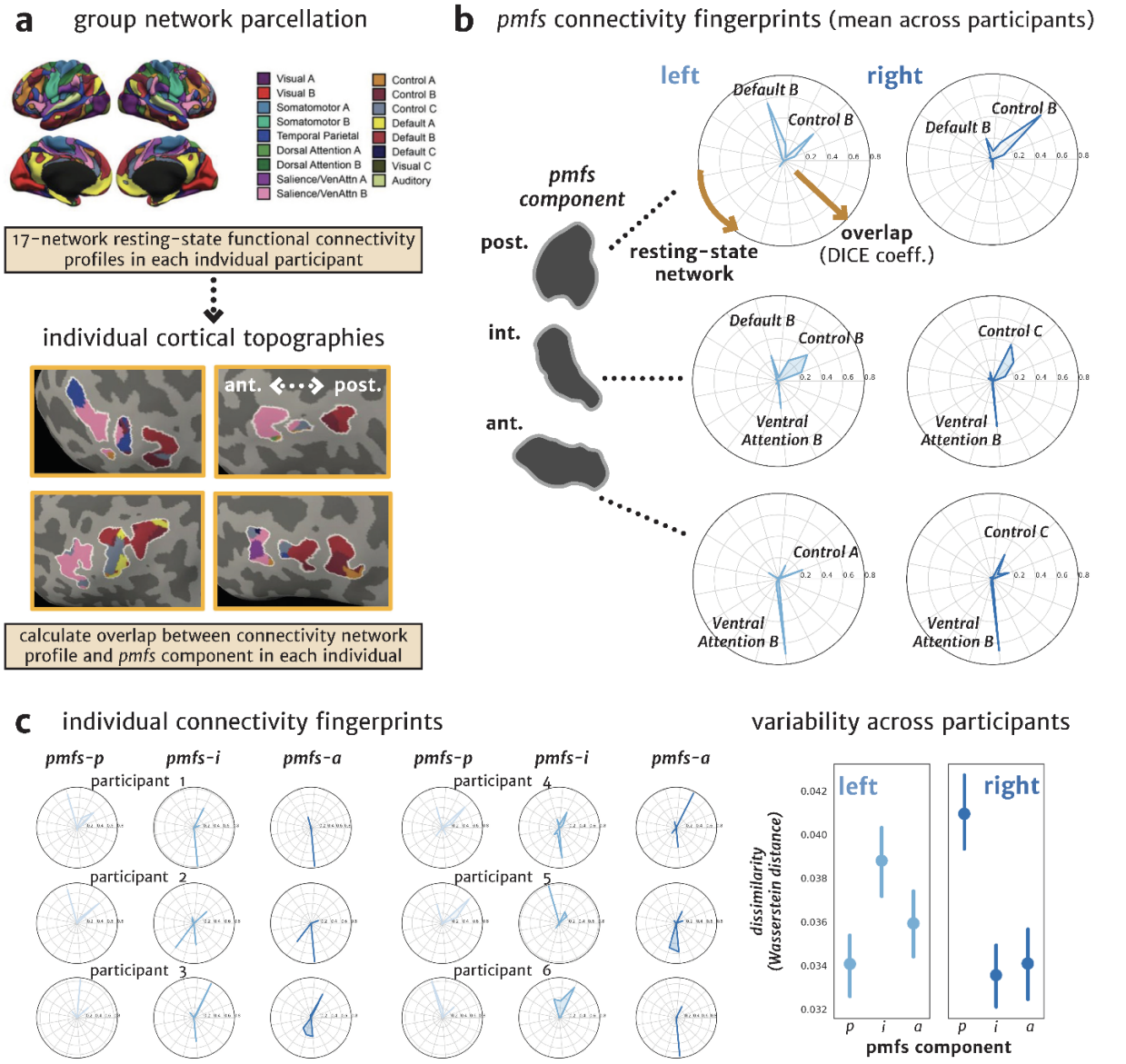


Figure 5. The *pmfs* components are functionally differentiable based on connectivity fingerprints within individuals. (**a**) Schematic of how individual-level resting state connectivity profiles were generated in each participant. Resting-state network parcellations for each participant were obtained from a recent study (Kong et al., 2018) in an observer-independent fashion of sulcal definitions in LPFC. Example individual cortical topographies are shown in four individual participants, colored according to the group parcellation. The individual cortical topographies and *pmfs* sulcal definitions were used to calculate the connectivity fingerprint, which represents the overlap of each network within the *pmfs* component of each participant. (**b**) Polar plots showing the mean connectivity fingerprint of the three *pmfs* components (plotted outwards) with each of 17 resting-state functional connectivity networks, across participants. Resting-state networks with the highest overlap across participants are labeled. (**c**) Left: Polar plots showing variability among 6 individual participants. Right: Dissimilarity of the resting-state network fingerprints (variability in the connectivity fingerprint across participants represented by the Wasserstein distance between unique pairs of participants; **Materials and Methods**) are plotted as a function of each *pmfs* component for left and right hemispheres. Error bars represent 68% CI (SEM) across unique participant pairs.

Our approach demonstrated that the *pmfs-p*, *pmfs-i*, and *pmfs-a* have different connectivity fingerprints and thus, are functionally dissociable. Average connectivity fingerprints across participants are illustrated in **Figure 5b**. A repeated-measures ANOVA with sulcal component (*pmfs-p, pmfs-i, pmfs-a*), hemisphere (left, right), and network yielded a significant component x network interaction ($F(32, 1120) = 45.2$, $p < 0.001$, $= 0.29$), as well as a component x network x hemisphere interaction ($F(32, 1120) = 5.26$, $p < 0.001$, $= 0.040$) (**Figure 5b**). In each hemisphere, there is a component x network interaction (left: $F(32, 1120) = 29.4$, $p < 0.001$, $= 0.35$, right: $F(32, 1120) = 23.2$, $p < 0.001$, $= 0.27$) in which the difference between hemispheres is driven by the *pmfs-p* connectivity fingerprint. Specifically, the *pmfs-p* overlaps most with the default mode network in the left hemisphere and the cognitive control network in the right hemisphere.

Additionally, there are also individual and hemispheric differences in the connectivity fingerprint of each *pmfs* component at the level of individual participants (**Figure 5c**; **Extended Data Figure 5-1**). To characterize individual differences, we built on work showing network connectivity variations across individuals (Kong et al., 2018; Seitzman et al., 2019) by relating this connectivity variability to individual anatomical landmarks in LPFC. We quantified connectivity fingerprint variability by measuring the pairwise Wasserstein distance between the connectivity profiles for all unique participant pairs for each sulcal component, in which a larger distance indicates decreased similarity, and therefore greater variability (see **Materials and Methods**). This approach quantifies how variable the pattern of network overlap (connectivity fingerprint) is across individuals for each *pmfs* component (**Figure 5c**, right). In the right hemisphere, the *pmfs-p* showed the most variable network profile across all unique participant pairs (*pmfs-p* vs. *pmfs-i,* Wilcoxson-Signed rank test, $W = 7.2\text{x}10^4$, $p < 0.001$, *pmfs-p* vs. *pmfs-a,* $W = 7.4\text{x}10^4$, $p < 0.001$), while the *pmfs-i* was most variable in the left hemisphere (*pmfs-i* vs. *pmfs-a,* $W = 8.8\text{x}10^4$, $p = 0.014$, *pmfs-i* vs. *pmfs-p,* $W = 8.0\text{x}10^4$, $p < 0.001$). This analysis suggests that the right *pmfs-p* and left *pmfs-i* mark regions of LPFC with particularly high levels of individual differences in functional connectivity profiles, providing an anatomical substrate for network connectivity differences across individuals.

The pmfs-p, pmfs-i, and pmfs-a are functionally dissociable: Meta-analyses across 83 experimental task categories

We next tested if the dissociation of functional networks between the *pmfs-p*, *pmfs-i*, and *pmfs-a* identified in individual participants (**Figure 5**) can also be observed in meta-analytic analyses of functional activation data at the group-level. That is, do the components of the *pmfs* show a functional dissociation of engagement over a wide array of cognitive operations? To test for different patterns of functional activations across tasks, we generated sulcal probability maps on a template cortical surface (**Figure 6a**, bottom left). Analogous to probabilistic maps for functional regions (Wang et al., 2015; Weiner et al., 2017; Weiner et al., 2018), the maps provide a vertex-wise measure of anatomical overlap across individuals for all 13 LPFC sulci examined in the present study. As the *pmfs* components disappear on average templates (**Figure 1**), these probabilistic maps are independent of the sulcal patterning of the template itself, which merely serves as a cortical surface independent of each individual cortical surface. We then compared these sulcal probability maps to 14 probabilistic "cognitive component" maps derived from an author-topic model of meta-analytic activation data across 83 experimental task categories (Yeo et al., 2015).

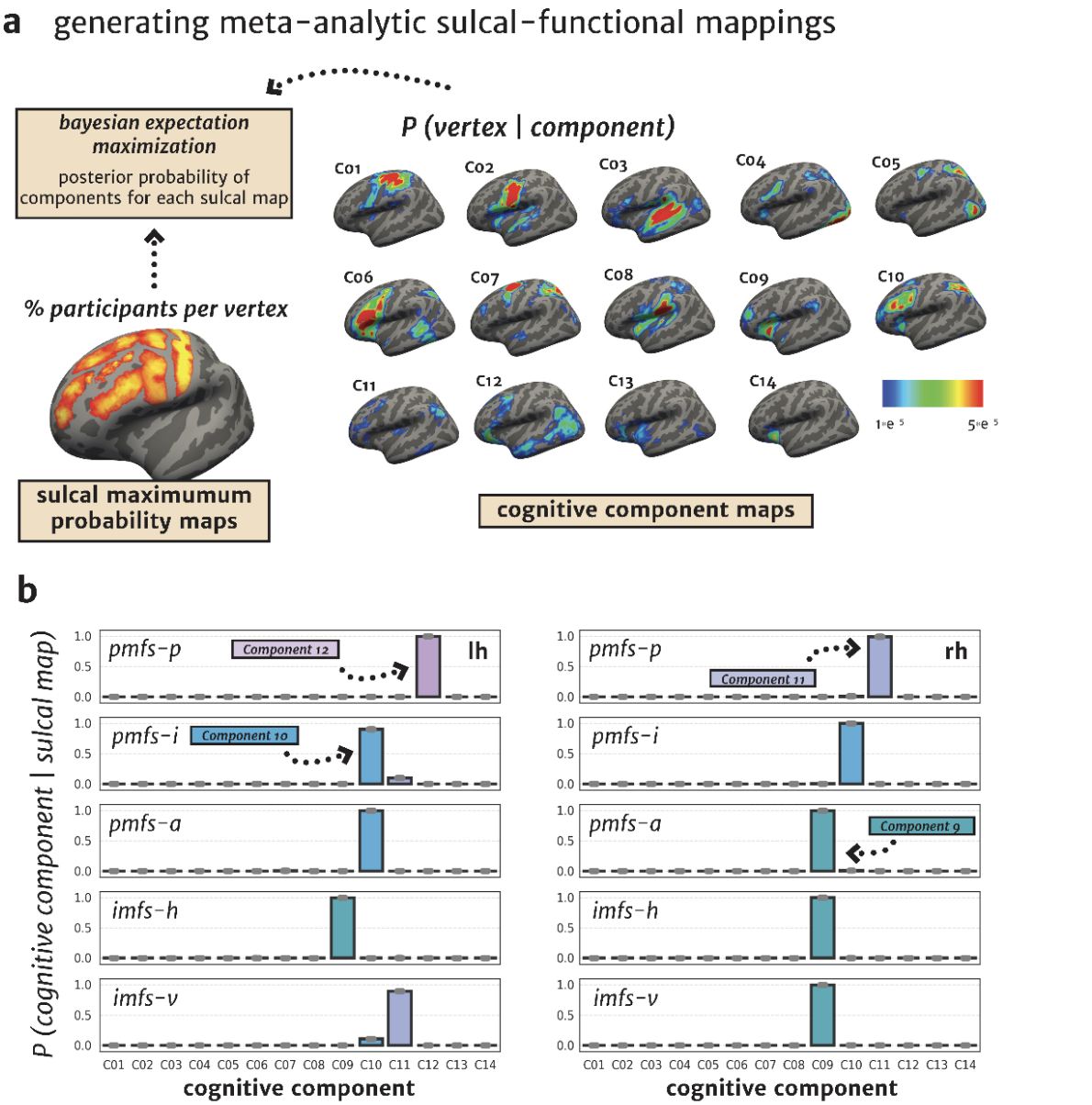


Figure 6. The *pmfs* and *imfs* components are functionally differentiable based on cognitive components: A meta-analysis of fMRI experimental tasks. (**a**) Schematic of analyses linking sulcal probability maps (bottom, left) and cognitive component maps (right) from a meta-analysis of fMRI experimental tasks (Yeo et al., 2015) using an expectation maximization algorithm (**Materials and Methods**). For each *pmfs* component, the algorithm provides a posterior probability for each of 14 cognitive components being associated with the provided sulcal probability map. (**b**) For each *pmfs* and *imfs* component in each hemisphere, the posterior probability for each cognitive component is plotted. This approach further supports that the *pmfs-p* (Component 12, lh; Component 11, rh), *pmfs-i* (Component 10, lh and rh), and *pmfs-a* (Component 10, lh; Component 9, rh; **Materials and Methods**) are functionally dissociable based on meta-analytic data of cognitive task activations. The *imfs-h* and *imfs-v* are also dissociable from the *pmfs* components in the left hemisphere, and functionally similar to the *pmfs-a* in the right hemisphere. Gray dots indicate individual participant data points when the analysis is performed with individual labels transformed to a template cortical surface, rather than with probability maps (**Materials and Methods**).

The cognitive component model links patterns of brain activity to behavioral tasks via latent components representing putative functional subsystems (Yeo et al., 2015). Each cognitive component map (which was calculated on the same template cortical surface used here) provides the probability that a given voxel will be activated by each of the 14 components (across all 83 tasks). We then used an expectation maximization algorithm (via posterior probability, **Materials and Methods**) to relate brain activity in each sulcal probability map to each cognitive component (**Figure 6a**, right). Importantly, when calculating the posterior probabilities, we implemented a leave-one-participant-out cross-validation procedure when constructing the sulcal probability maps in order to assess variability in the generated posterior probabilities for each cognitive component (**Figure 6b**). To indicate feasibility of this approach, the somato-motor components of the cognitive component map (C01, C02) align most highly with the central sulcus as one would expect, which shows the ability of this method to measure structural-functional correspondences at the meta-analytic level.

This approach further reveals that the *pmfs-p*, *pmfs-i*, and *pmfs-a* are functionally dissociable based on meta-analytic data of cognitive task activations. In the right hemisphere, the *pmfs-p*, *pmfs-i*, and *pmfs-a* showed distinct probabilities for separate cognitive components: 1) the *pmfs-p* loaded onto a default mode component (C11), 2) the *pmfs-i* loaded onto an executive function component (C10), and 3) the *pmfs-a* loaded onto an inhibitory control component (C09). In the left hemisphere, the *pmfs-a* and *pmfs-i* both loaded onto an executive function (C10) component, while the *pmfs-p* loaded onto an emotional processing/episodic memory component (C12). The *pmfs* was also dissociable in activation profiles from the more anterior *imfs*. In the left hemisphere, the imfs showed no overlap with the *pmfs*, with the *imfs-h* loading onto the inhibitory control component (C09), and the *imfs-v* loading onto a default mode component (C11). In the right hemisphere, both the *imfs-h and imfs-v* loaded onto the same inhibitory control component (C09) as the *pmfs-a*.

Like our individual participant analyses, there were also hemispheric differences: the cognitive components overlapping the most with the *pmfs-a* and *pmfs-p* differed between the two hemispheres. The *pmfs-p* loaded onto an emotional processing/episodic memory component in the left hemisphere (**Figure 6b**, top row) and a default mode component in the right hemisphere (**Figure 6b**, top row), while the *pmfs-a* loaded onto an executive function component in the left hemisphere (**Figure 6b**, third row) and an inhibitory control component in the right hemisphere (**Figure 6b**, third row).

Finally, previous studies have identified retinotopic representations in human LPFC (Hagler and Sereno, 2006; Kastner et al., 2007; Mackey et al., 2017), but the three *pmfs* components did not overlap with cognitive components associated with visual processing in these meta-analytic analyses. To further examine the relationship between the *pmfs* components and visual processing, we analyzed whether the *pmfs* components explained a significant amount of variance (**Figure 7**) in a newly published, whole brain dataset of population receptive field measurements in 181 participants (Benson et al., 2018). When considering voxels that demonstrate retinotopic responses ($R^2 > 15\%$), the highest overlap between predicted *pmfs* location and retinotopic representations was specific to the right hemisphere for the *pmfs-i* (mean R^2 across participants = 28.5%), with less overlap in the left hemisphere (all other *pmfs* R^2 values < 20%). The most consistent correspondence between visual field maps and sulcal location occurred at (1) the intersection of

the *sprs* and *sfs-p*, and (2) the intersection of the *iprs* and *ifs*, as previously reported ((Mackey et al., 2017); **Figure 7**). The *iprs* showed the highest retinotopic responses of the LPFC sulci (lh: 34.2%; rh: 48.9%) measured here, and this is also consistent with a recent study identifying a region critical for conditional eye movements within a similar location in the *ifs* (Germann and Petrides, 2020). Future studies examining the relationship between *pmfs* components and retinotopic representations in individual participants will further expand on these findings.

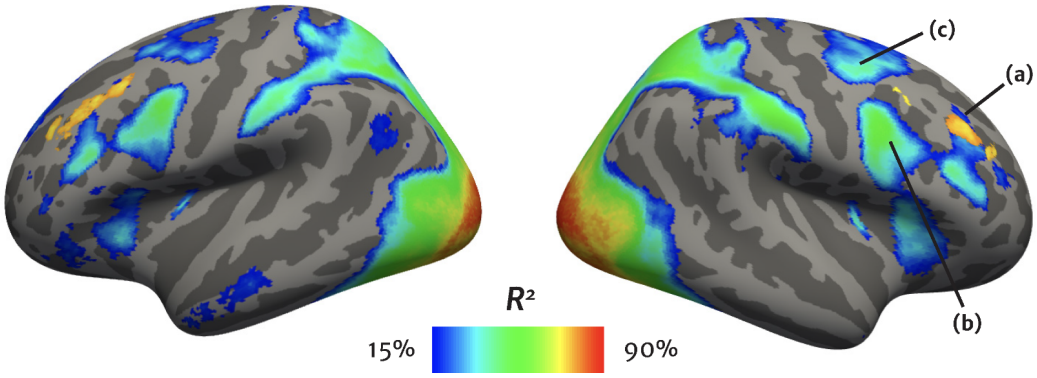


Figure 7. Comparing the overlap between retinotopic responses relative to the predicted location of the *pmfs* sulcal components. Map of the mean (n = 181) R^2 metric (colorbar) from the HCP retinotopy dataset (Benson et al., 2018) on the *fsaverage* template cortical surface for each hemisphere, thresholded at 15%. This metric measures how well the fMRI time-series at each vertex is modeled by population receptive field (pRF) modeling that was calculated and shared by Benson and colleagues (https://osf.io/bw9ec/wiki/home/). Predicted *pmfs* location from the maximum probability maps is overlaid in orange (thresholded at 33% overlap across participants). There was only a modest overlap between predicted *pmfs* location and retinotopic representations (a) in the right hemisphere (no overlap in the left hemisphere). Instead, and consistent with prior work (Mackey et al., 2017), the highest correspondence between retinotopic responses and sulcal patterning in LPFC occurs at two sulcal intersections: 1) the *sprs* and *sfs-p* (c), and (2) the *iprs* and *ifs* (b).

Extensive individual differences in the location of the pmfs across individuals

Although the three *pmfs* components are prominent within each hemisphere, there is extensive individual variability in the precise location of each sulcal component within the posterior MFG. To determine how well the probability maps could predict the location of the *pmfs-p*, *pmfs-i*, and *pmfs-a* within *individual* hemispheres, we used a cross-validated approach, iteratively leaving out one participant from the calculation of probability maps (**Figure 8a**). Then, the maximum probability maps (MPMs) were projected to the held-out individual's native cortical surface to calculate the overlap between the manually identified and probabilistically identified sulcal locations. This procedure resulted in a measure of location variability for each sulcal component (**Figure 8b**). For these calculations, we used the *central sulcus* (*cs*) as a noise ceiling (left: $cs = 0.85 \pm 0.02$; right: $cs = 0.85 \pm 0.06$) as it is a) considered very stable across individuals (see **Materials and Methods**) and b) used in the cortex-based alignment procedure (Fischl et al., 1999b).

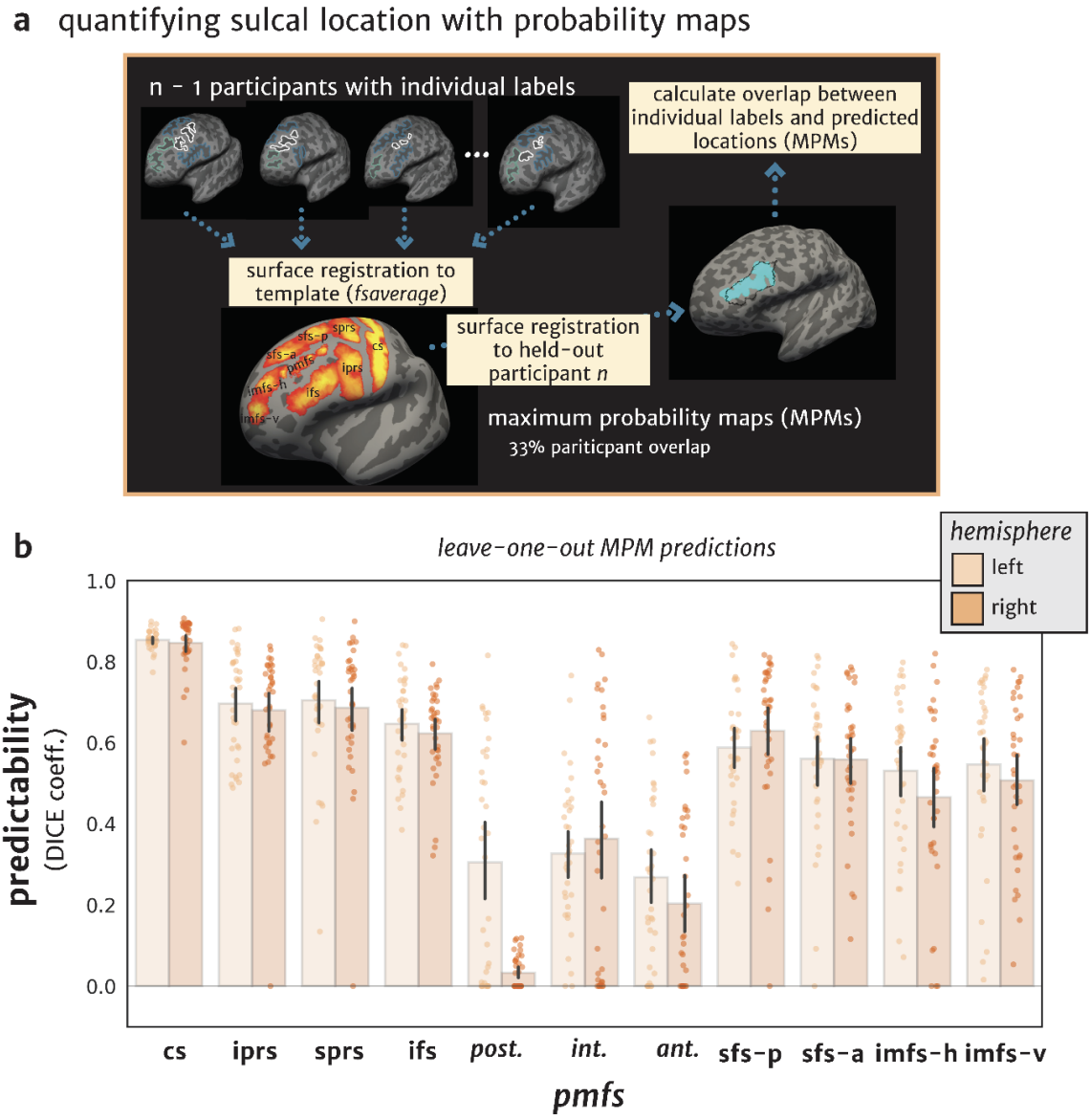


Figure 8. Quantification and prediction of *pmfs-p, pmfs-i, and pmfs-a* within individual hemispheres. (**a**) Procedure to generate sulcal probability maps based on the manual anatomical labeling within each individual participant. Labels from each individual are transformed to a template cortical surface to form a probabilistic sulcal map and then projected onto the surface of a held-out individual participant. The overlap between the manual anatomical label on the held-out participant and predicted location was then calculated for each iteration across participants. (**b**) Overlap (DICE coefficient) between predicted and manual location of each *pmfs* component within individual participants. Prediction for the *pmfs* is highest when all three components are combined. The central sulcus (*cs*) is included as a noise ceiling for reference, as this landmark is used in the surface registration algorithm that aligns cortical surfaces across participants.

The *pmfs* components exhibited significant variability in sulcal location across participants (left: *pmfs-p* = 0.30 ± 0.28, *pmfs-i* = 0.32 ± 0.18, *pmfs-a* = 0.27 ± 0.20; right: *pmfs-p* = 0.03 ± 0.04, *pmfs-i* = 0.37 ± 0.18, *pmfs-a* = 0.20 ± 0.20). A 2-way repeated-measures ANOVA with *pmfs* sulcal component (*pmfs-p, pms-i, pmfs-a*) and hemisphere (*right*, *left*) revealed a sulcus x hemisphere interaction ($F(1.84, 64.47) = 9.52$, $p < 0.001$, $= 0.08$) driven by the finding that the *pmfs-p* is highly variable across individuals, resulting in very little predictability in the right hemisphere (**Figure 8b**). When using all three *pmfs* components together, prediction is more robust (left: *pmfs* = 0.41 ± 0.13; right: *pmfs* = 0.37 ± 0.15), but still much lower than the predictability of the *cs* and also lower than prediction performance for all other LPFC sulci quantified in the present study (**Figure 8b**). These results demonstrate that although the *pmfs* is prominent within each individual (**Extended Data Figure 2-1**), the location of each *pmfs* component is variable across individuals, which provides empirical support for the historical confusion regarding its identification and labeling (**Figure 1**).

Discussion

Here, we examined the relationship between cortical anatomy and function in human lateral prefrontal cortex (LPFC) and showed for the first time (to our knowledge) that the posterior middle frontal sulcus (*pmfs*) serves as a meso-scale link between myelin content and functional connectivity in individual participants. The *pmfs* is a characteristically shallow tertiary sulcus with three components that differ in their myelin content, resting state connectivity profiles, and engagement across meta-analyses of 83 cognitive tasks. We first discuss how these findings suggest modern empirical support for a classic, yet largely unconsidered, anatomical theory (Sanides, 1962, 1964), as well as a recent cognitive neuroscience theory proposing a functional hierarchy in LPFC (Koechlin and Summerfield, 2007; Badre and D'Esposito, 2009; Badre and Nee, 2018). We end by discussing a growing need for computational tools that automatically define tertiary sulci throughout cortex.

The anatomical-functional coupling in LPFC identified here is quite surprising considering the widespread literature providing little support for fine-grained anatomical-functional coupling in this cortical expanse and in association cortices more broadly when conducting traditional group-analyses (Paquola et al., 2019; Vazquez-Rodriguez et al., 2019). Indeed, cortical folding patterns relative to the location of anatomical, functional, or multimodal transitions are considered "imperfectly correlated" (Welker, 1990; Glasser et al., 2016) in association cortices and especially in LPFC (Van Essen et al., 2012; Caspers et al., 2013; Robinson et al., 2014; Coalson et al., 2018). Contrary to these previous findings that did not consider tertiary sulci, the present findings appear to support a classic, yet largely unconsidered theory proposed by Sanides (1962, 1964) that tertiary sulci are potentially meaningful anatomical and functional landmarks in association cortices – and in particular, in LPFC. Specifically, Sanides proposed that because tertiary sulci emerge late in gestation and exhibit a protracted postnatal development, they likely serve as functional and architectonic landmarks in human association cortices, which also exhibit a protracted postnatal development. Sanides (1964) further proposed that the late morphological development of tertiary sulci is likely related to protracted cognitive skills associated with LPFC. Interestingly, identifying *pmfs* components in his classic images shows myeloarchitectonic gradations among five areas in LPFC (**Figure 9a**). Linking these data to recent modern parcellations of the human cerebral cortex (Sallet et al., 2013; Glasser et al., 2016) shows that *pmfs* components likely serve as boundaries

among a series of cortical areas, which can be addressed in future research in individual participants (**Figure 9b**).

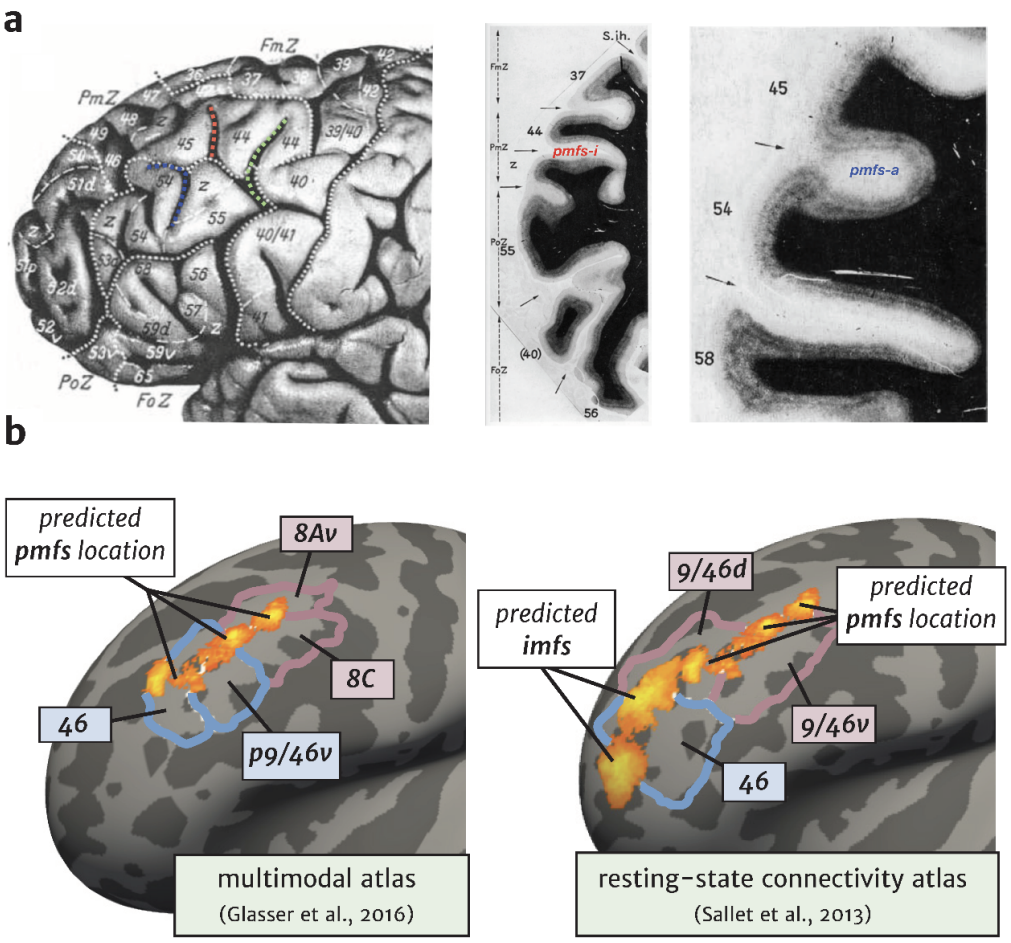


Figure 9. Linking the past to the present: Myelination gradients, cortical areas, and the *pmfs*. (**a**) Left: Photograph of a left hemisphere from Sanides (1962). Numbers indicate cortical areas differing in myeloarchitecture. Dotted white lines: Sulcal boundaries as defined by Sanides. Dotted colored lines: *pmfs-p* (green), *pmfs-i* (red), and *pmfs-a* (blue) based on modern definitions used in the present study. Identifying *pmfs* components in Sanides' classic images shows that he identified myeloarchitectonic gradations within *pmfs* components, which is consistent with the present measurements. Gradations occurred in superior-inferior, as well as anterior-posterior dimensions. In the inferior portion of the *pmfs-p* (green), there is an anterior-posterior transition between areas 40 and 55. In the *pmfs-i* (red), there are two transitions: (i) a superior-inferior transition between areas 44 and a transition zone to area 55, and (ii) an anterior-posterior transition between areas 44 and 45. In the *pmfs-a*, there is a transition between areas 45 and 54. Right: Myelination stain of a histological section (coronal orientation) from Sanides (1962). Arrows indicate boundaries between labeled myeloarchitectonic areas (numbers). *pmfs-a* is labeled to help the reader link the myelination stain to the image at left. The reader can appreciate the shallowness of the *pmfs-a* relative to the sulcus (*ifs*) between areas 54 and 58, which is also consistent with our measurements (**Figure 2**). (**b**) Left: Maximum probability maps (thresholded at 33% overlap across participants) for the *pmfs-p*, *pmfs-i*, and *pmfs-a* are shown on the FreeSurfer average template (left hemisphere). The probability maps are shown relative to four areas from a multi-modal cortical parcellation based on structural and functional MRI data (Glasser et al., 2016). The *pmfs-a* appears to denote the dorsal to ventral transition between areas 46 and p9/46v in anterior LPFC, while the *pmfs-p* appears to denote the dorsal to ventral transition between areas 8Av and 8C in posterior LPFC. Right: *pmfs* and *imfs* maximum

probability maps relative to a resting-state fMRI parcellation with proposed homologous parcels between monkey and human LPFC from Sallet et al., 2013. Here, the *pmfs-i and pmfs-a* denote the 9/46d and 9/46v boundary, while the *imfs* is situated within area 46. This relationship is also consistent with a recent cytoarchitectonic atlas showing that the *pmfs-a* identifies a transition between 9/46v and 9/46d (Petrides, 2019).

In addition to supporting Sanides' classic anatomical theory, the present data demonstrated that the three *pmfs* components exhibit different resting-state connectivity profiles along a rostral-caudal axis, which builds on previous work also supporting a functional hierarchy along a rostral-caudal axis of LPFC. Further consistent with this hierarchy, evidence from neuroimaging, lesion, and electrocorticography studies indicate that this proposed rostral-caudal axis of LPFC is also related to levels of temporal and cognitive abstraction. That is, more anterior LPFC cortical regions are more highly engaged in tasks with higher abstract complexity (Koechlin et al., 2003; Koechlin and Summerfield, 2007; Voytek et al., 2015; Mansouri et al., 2017). While there is axonal tracing data in non-human primates suggesting an anatomical basis for such a hierarchical organization (Goulas et al., 2014; Goulas et al., 2019), the present findings provide new evidence for anatomically and functionally dissociable sulcal components in LPFC that also support a hierarchical organization within individuals. Future work leveraging finer-scale multimodal and microanatomical data from individual human brains will be critical for uncovering anatomical and functional properties of LPFC across spatial and temporal scales that may further support the proposed functional rostral-caudal hierarchy of human LPFC.

Together, the culmination of present and previous findings suggest that tertiary sulci are landmarks in human ventral temporal cortex (Nasr et al., 2011; Caspers et al., 2013; Weiner et al., 2014; Lorenz et al., 2017), medial PFC (Amiez et al., 2019; Lopez-Persem et al., 2019), and now, LPFC. This begs the question: How many other tertiary sulci serve as cortical landmarks? We stress that it is unlikely that all tertiary sulci will serve as cortical landmarks, since neuroanatomists have known for over a century that not all sulci function as cortical landmarks (Smith, 1907; Bailey and Bonin, 1951; Ono et al., 1990; Welker, 1990; Van Essen et al., 2019). Nonetheless, this does not preclude the importance of future studies identifying which tertiary sulci are architectonic, functional, behavioral, or multimodal landmarks – not only in healthy young adults as examined here, but also in developmental (Voorhies et al., 2020) and clinical (Garrison et al., 2015; Brun et al., 2016) cohorts. Additionally, tertiary sulci can also serve as evolutionary markers for primate cortical homology. For example, shallow "dimples" co-occur with the frontal eye field (FEF) in macaques, while deeper sulci co-occur with the proposed homologue of the FEF in humans (Amiez and Petrides, 2009; Schall et al., 2020). Humans may also have tertiary sulci in locations that non-human primates do not have dimples as was recently shown in medial PFC (Amiez et al., 2019).

Carefully examining the relationship among tertiary sulci and multiple types of anatomical, functional, and behavioral data in individual participants will require new neuroimaging tools to automatically identify tertiary sulci throughout human cortex. For instance, most neuroimaging software packages are only capable of automatically defining ~30-35 primary and secondary sulci in a given hemisphere (Destrieux et al., 2010). Current estimates approximate ~110 sulci in each hemisphere when considering tertiary sulci (Petrides, 2019). Thus, studies in the immediate future will still require the manual identification of tertiary sulci, which is labor intensive and requires expertise ((Miller et al., 2020a) for a historical discussion regarding the manual labeling of tertiary sulci in LPFC). For example, the present study required manual definitions of 936 sulci in 72 hemispheres. While 72 is a large sample size compared to other labor-intensive anatomical studies in which 20 hemispheres is considered sufficient to encapsulate individual differences (Amunts

and Zilles, 2015; Amunts et al., 2020), 2400 hemispheres are available just from the HCP alone. Defining tertiary sulci in only the LPFC of every HCP participant would require ~26,400 manual definitions, while defining all tertiary sulci in the entire HCP dataset would require over a quarter of a million (~256,800) manual definitions. Consequently, manual identification of tertiary sulci will continue to limit sample sizes in immediate future studies until new automated methods are generated (Klein et al., 2017; Hao et al., 2020; Lyu et al., 2020).

In the interim, we sought to leverage the anatomical labeling in this study to aid the field in the identification of sulcal landmarks in LPFC. The probability maps of sulcal locations in the present study are openly available and may be transformed to held-out individual brains (**Figure 9**). Accordingly, manual identification of these landmarks within individuals is greatly aided, allowing future studies to apply these tools to identify LPFC tertiary in individual participants, including those from various groups such as patient or developmental cohorts. Because smaller tertiary sulci in association cortex are the latest sulcal indentations to develop (Sanides, 1962, 1964; Chi et al., 1977; Welker, 1990; Armstrong et al., 1995), their anatomical trajectories and properties likely relate to the development of cognitive abilities associated with the LPFC and other association areas as Sanides hypothesized, which recent ongoing work supports (Voorhies et al., 2020). Moving forward, we hope to leverage the manual labeling performed here to develop better automated algorithms for sulcal labeling within individuals. Future work using deep learning algorithms may help to identify tertiary structures in novel brains without manual labeling or intervention (Borne et al., 2020; Hao et al., 2020; Lyu et al., 2020). Such automated tools have translational applications as tertiary sulci are largely hominoid-specific structures (Amiez et al., 2019; Miller et al., 2020b) located in association cortices associated with pathology in many neurological disorders. Thus, morphological features of these under-studied neuroanatomical structures may be useful clinical biomarkers for future diagnostic purposes. To begin to achieve this goal and to aid the field, we share our probabilistic maps of LPFC tertiary sulci with the publication of this paper.

References

Amiez C, Petrides M (2007) Selective involvement of the mid-dorsolateral prefrontal cortex in the coding of the serial order of visual stimuli in working memory. Proc Natl Acad Sci U S A 104:13786-13791.

Amiez C, Petrides M (2009) Anatomical organization of the eye fields in the human and non-human primate frontal cortex. Prog Neurobiol 89:220-230.

Amiez C, Sallet J, Hopkins WD, Meguerditchian A, Hadj-Bouziane F, Ben Hamed S, Wilson CRE, Procyk E, Petrides M (2019) Sulcal organization in the medial frontal cortex provides insights into primate brain evolution. Nat Commun 10:3437.

Amunts K, Zilles K (2015) Architectonic Mapping of the Human Brain beyond Brodmann. Neuron 88:1086-1107.

Amunts K, Mohlberg H, Bludau S, Zilles K (2020) Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. Science 369:988-992.

Annese J, Pitiot A, Dinov ID, Toga AW (2004) A myelo-architectonic method for the structural classification of cortical areas. Neuroimage 21:15-26.

Armstrong E, Schleicher A, Heyder O, Maria C, Zilles K (1995) The Ontogeny of Human Gyrification. Cereb Cortex 5.

Badre D, D'Esposito M (2009) Is the rostro-caudal axis of the frontal lobe hierarchical? Nat Rev Neurosci 10:659-669.

Badre D, Nee DE (2018) Frontal Cortex and the Hierarchical Control of Behavior. Trends Cogn Sci 22:170-188.

Bailey P, Bonin GV (1951) The Isocortex of Man. Urbana: University of Illinois Press.

Barrett RLC, Dawson M, Dyrby TB, Krug K, Ptito M, D'Arceuil H, Croxson PL, Johnson PJ, Howells H, Forkel SJ, Dell'Acqua F, Catani M (2020) Differences in Frontal Network Anatomy Across Primate Species. J Neurosci 40:2094-2107.

Benson NC, Jamison KW, Arcaro MJ, Vu AT, Glasser MF, Coalson TS, Van Essen DC, Yacoub E, Ugurbil K, Winawer J, Kay K (2018) The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. J Vis 18:23.

Borne L, Riviere D, Mancip M, Mangin JF (2020) Automatic labeling of cortical sulci using patch- or CNN-based segmentation techniques combined with bottom-up geometric constraints. Med Image Anal 62:101651.

Braitenberg V (1962) A note on myeloarchitectonics. J Comp Neurol 118:141-156.

Brun L, Auzias G, Viellard M, Villeneuve N, Girard N, Poinso F, Da Fonseca D, Deruelle C (2016) Localized Misfolding Within Broca's Area as a Distinctive Feature of Autistic Disorder. Biol Psychiatry Cogn Neurosci Neuroimaging 1:160-168.

Caspers J, Zilles K, Eickhoff SB, Schleicher A, Mohlberg H, Amunts K (2013) Cytoarchitectonical analysis and probabilistic mapping of two extrastriate areas of the human posterior fusiform gyrus. Brain Struct Funct 218:511-526.

Chi JG, Dooling EC, Gilles FH (1977) Gyral development of the human brain. Ann Neurol 1:86-93.

Coalson TS, Van Essen DC, Glasser MF (2018) The impact of traditional neuroimaging methods on the spatial localization of cortical areas. Proc Natl Acad Sci U S A 115:E6356-E6365.

Connolly C (1950) External Morphology of the Primate Brain. Springfield.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179-194.

Demirtas M, Burt JB, Helmer M, Ji JL, Adkinson BD, Glasser MF, Van Essen DC, Sotiropoulos SN, Anticevic A, Murray JD (2019) Hierarchical Heterogeneity across Human Cortex Shapes Large-Scale Neural Dynamics. Neuron 101:1181-1194 e1113.

Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53:1-15.

Dick F, Tierney AT, Lutti A, Josephs O, Sereno MI, Weiskopf N (2012) In vivo functional and myeloarchitectonic mapping of human primary auditory areas. J Neurosci 32:16095-16105.

Ding SL et al. (2016) Comprehensive cellular-resolution atlas of the adult human brain. J Comp Neurol 524:3127-3481.
Donahue CJ, Glasser MF, Preuss TM, Rilling JK, Van Essen DC (2018) Quantitative assessment of prefrontal cortex in humans relative to nonhuman primates. Proc Natl Acad Sci U S A.
Eberstaller O (1890) Das Stirnhirn. Wien: Urban & Schwarzenberg.
Fischl B, Sereno MI, Dale AM (1999a) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195-207.
Fischl B, Sereno MI, Tootell RBH, Dale AM (1999b) High-Resolution Intersubject Averaging and a Coordinate System for the Cortical Surface. Human Brain Mapping 8:272-284.
Fisher AV (2019) Selective sustained attention: a developmental foundation for cognition. Current Opinion in Psychology 29:248-253.
Flechsig P (1920) Anatomie des Menschlichen Gehirns und Rückenmarks auf Myelogenetischer Grundlage. Leipzig.
Fritz CO, Morris PE, Richler JJ (2012) Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 141:2-18.
Gao JS, Huth AG, Lescroart MD, Gallant JL (2015) Pycortex: an interactive surface visualizer for fMRI. Front Neuroinform 9:23.
Garrison JR, Fernyhough C, McCarthy-Jones S, Haggard M, Australian Schizophrenia Research B, Simons JS (2015) Paracingulate sulcus morphology is associated with hallucinations in the human brain. Nat Commun 6:8956.
Germann J, Petrides M (2020) Area 8A within the Posterior Middle Frontal Gyrus Underlies Cognitive Selection between Competing Visual Targets. eNeuro 7.
Glasser MF, Van Essen DC (2011) Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. J Neurosci 31:11597-11616.
Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, Van Essen DC (2016) A multi-modal parcellation of human cerebral cortex. Nature 536:171-178.
Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, Van Essen DC, Jenkinson M, Consortium WU-MH (2013) The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80:105-124.
Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C, Gratton C, Sun H, Hampton JM, Coalson RS, Nguyen AL, McDermott KB, Shimony JS, Snyder AZ, Schlaggar BL, Petersen SE, Nelson SM, Dosenbach NUF (2017) Precision Functional Mapping of Individual Human Brains. Neuron.
Goulas A, Uylings HB, Stiers P (2014) Mapping the hierarchical layout of the structural network of the macaque prefrontal cortex. Cereb Cortex 24:1178-1194.
Goulas A, Majka P, Rosa MGP, Hilgetag CC (2019) A blueprint of mammalian cortical connectomes. PLoS Biol 17:e2005346.
Hagler DJ, Jr., Sereno MI (2006) Spatial maps in frontal and prefrontal cortex. Neuroimage 29:567-577.
Hao L, Bao S, Tang Y, Gao R, Parvathaneni P, Miller JA, Voorhies W, Yao J, Bunge SA, Weiner KS, Landman BA, Lyu I (2020) Automatic Labeling of Cortical Sulci using Convolutional Neural Networks in a Developmental Cohort. In: IEEE International Symposium on Biomedical Imaging (ISBI). Iowa City, IA.
Hinds OP, Rajendran N, Polimeni JR, Augustinack JC, Wiggins G, Wald LL, Diana Rosas H, Potthast A, Schwartz EL, Fischl B (2008) Accurate prediction of V1 location from cortical folds in a surface coordinate system. Neuroimage 39:1585-1599.
Hopf A (1956) Über die Verteilung myeloarchitektonischer Merkmale in der Stirnhirnrinde beim Menschen. J Hirnforsch 2:311-333.
Kastner S, DeSimone K, Konen CS, Szczepanski SM, Weiner KS, Schneider KA (2007) Topographic maps in human frontal cortex revealed in memory-guided saccade and spatial working-memory tasks. J Neurophysiol 97:3494-3507.

Klein A, Ghosh SS, Bao FS, Giard J, Hame Y, Stavsky E, Lee N, Rossa B, Reuter M, Chaibub Neto E, Keshavan A (2017) Mindboggling morphometry of human brains. PLoS Comput Biol 13:e1005350.
Koechlin E, Summerfield C (2007) An information theoretical approach to prefrontal executive function. Trends Cogn Sci 11:229-235.
Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. Science 302:1181-1185.
Kong R, Li J, Orban C, Sabuncu MR, Liu H, Schaefer A, Sun N, Zuo XN, Holmes AJ, Eickhoff SB, Yeo BTT (2018) Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. Cereb Cortex.
Lopez-Persem A, Verhagen L, Amiez C, Petrides M, Sallet J (2019) The Human Ventromedial Prefrontal Cortex: Sulcal Morphology and Its Influence on Functional Organization. J Neurosci 39:3627-3639.
Lorenz S, Weiner KS, Caspers J, Mohlberg H, Schleicher A, Bludau S, Eickhoff SB, Grill-Spector K, Zilles K, Amunts K (2017) Two New Cytoarchitectonic Areas on the Human Mid-Fusiform Gyrus. Cereb Cortex 27:373-385.
Lyu I, Bao S, Hao L, Yao J, Miller JA, Voorhies W, Taylor WD, Bunge SA, Weiner KS, Landman BA (2020) Labeling Lateral Prefrontal Sulci using Spherical Data Augmentation and Context-aware Training. Neuroimage (in revision).
Mackey WE, Winawer J, Curtis CE (2017) Visual field map clusters in human frontoparietal cortex. Elife 6.
Madan CR (2019) Robust estimation of sulcal morphology. Brain Inform 6:5.
Mansouri FA, Koechlin E, Rosa MGP, Buckley MJ (2017) Managing competing goals - a key role for the frontopolar cortex. Nat Rev Neurosci 18:645-657.
Miller EK, Cohen JD (2001) AN INTEGRATIVE THEORY OF PREFRONTAL CORTEX FUNCTION. Annu Rev Neurosci 24:167-202.
Miller JA, D'Esposito M, Weiner KS (2020a) Using tertiary sulci to map the "cognitive globe" of prefrontal cortex. psyArxiv.
Miller JA, Voorhies WI, Li X, Raghuram I, Palomero-Gallagher N, Zilles K, Sherwood CC, Hopkins WD, Weiner KS (2020b) Sulcal morphology of ventral temporal cortex is shared between humans and other hominoids. Sci Rep 10:17132.
Nasr S, Liu N, Devaney KJ, Yue X, Rajimehr R, Ungerleider LG, Tootell RB (2011) Scene-selective cortical regions in human and nonhuman primates. J Neurosci 31:13771-13785.
Nee DE, D'Esposito M (2016) The hierarchical organization of the lateral prefrontal cortex. eLife 5.
Ono M, Kubik S, Abernathey C (1990) Atlas of the Cerebral Sulci. New York: Thieme Medical Publishers, Inc.
Paquola C, Vos De Wael R, Wagstyl K, Bethlehem RAI, Hong SJ, Seidlitz J, Bullmore ET, Evans AC, Misic B, Margulies DS, Smallwood J, Bernhardt BC (2019) Microstructural and functional gradients are increasingly dissociated in transmodal cortices. PLoS Biol 17:e3000284.
Petrides M (2019) Atlas of the Morphology of the Human Cerebral Cortex on the Average MNI Brain, 1 Edition. London, UK: Elsevier.
Petrides M, Pandya DN (2012) The-Frontal-Cortex. In: The Human Nervous System (Mai J, Paxinos G, eds), pp 988-1011: Elsevier.
Rajkowska G, Goldman-Rakic PS (1995) Cytoarchitectonic Definition of Prefrontal areas in the Normal Human Cortex: II. Variability in Locations of Areas 9 and 46 and Relationship to the Talairach Coordinate System. Cereb Cortex 5.
Retzius G (1896) Das Menschenhirn. Stockholm, Sweden: Norstedt and Soener.
Robinson EC, Jbabdi S, Glasser MF, Andersson J, Burgess GC, Harms MP, Smith SM, Van Essen DC, Jenkinson M (2014) MSM: a new flexible framework for Multimodal Surface Matching. Neuroimage 100:414-426.
Rollins CPE, Garrison JR, Arribas M, Seyedsalehi A, Li Z, Chan RCK, Yang J, Wang D, Lio P, Yan C, Yi ZH, Cachia A, Upthegrove R, Deakin B, Simons JS, Murray GK, Suckling J (2020) Evidence in cortical

folding patterns for prenatal predispositions to hallucinations in schizophrenia. Transl Psychiatry 10:387.
Rowley CD, Bazin PL, Tardif CL, Sehmbi M, Hashim E, Zaharieva N, Minuzzi L, Frey BN, Bock NA (2015) Assessing intracortical myelin in the living human brain using myelinated cortical thickness. Front Neurosci 9:396.
Sallet J, Mars RB, Noonan MP, Neubert FX, Jbabdi S, O'Reilly JX, Filippini N, Thomas AG, Rushworth MF (2013) The organization of dorsal frontal cortex in humans and macaques. J Neurosci 33:12255-12274.
Sanides F (1962) Die Architektonik Des Menschlichen Stirnhirns Zugleich Eine Darstellung Der Prin. In: Monographien aus dem Gesamtgebiete der Neurologie und Psychiatrie (Muller M, Spatz H, Vogel P, eds), pp 176-190. Berlin: Springer Berlin Heidelberg.
Sanides F (1964) STRUCTURE AND FUNCTION OF THE HUMAN FRONTAL LOBE. Neuropsychologia 2:209-219.
Sanides F (1972) Representation in the cerebral cortex and its areal lamination patterns. In: The Structure and Function of Nervous Tissue (Bourne GH, ed), pp 329-453. New York: Academic Press.
Schall JD, Zinke W, Cosman JD, Schall MS, Pare M, Pouget P (2020) On the Evolution of the Frontal Eye Field: Comparisons of Monkeys, Apes, and Humans. In: Evolutionary Neuroscience, 2 Edition (Kaas JH, ed), pp 861-883: Elsevier.
Seitzman BA, Gratton C, Laumann TO, Gordon EM, Adeyemo B, Dworetsky A, Kraus BT, Gilmore AW, Berg JJ, Ortega M, Nguyen A, Greene DJ, McDermott KB, Nelson SM, Lessov-Schlaggar CN, Schlaggar BL, Dosenbach NUF, Petersen SE (2019) Trait-like variants in human functional brain networks. Proc Natl Acad Sci U S A 116:22851-22861.
Semendeferi K, Lu A, Schenker N, Damasio H (2002) Humans and great apes share a large frontal cortex. Nat Neurosci 5:272-276.
Shams Z, Norris DG, Marques JP (2019) A comparison of in vivo MRI based cortical myelin mapping using T1w/T2w and R1 mapping at 3T. PLoS One 14:e0218089.
Smith GE (1907) A New Topographical Survey of the Human Cerebral Cortex, being an Account of the Distribution of the Anatomically Distinct Cortical Areas and their Relationship to the Cerebral Sulci. J Anat Physiol 41:237-254.
Szczepanski SM, Knight RT (2014) Insights into human behavior from lesions to the prefrontal cortex. Neuron 83:1002-1018.
Van Essen DC, Dierker DL (2007) Surface-based and probabilistic atlases of primate cerebral cortex. Neuron 56:209-225.
Van Essen DC, Glasser MF, Dierker DL, Harwell J (2012) Cortical parcellations of the macaque monkey analyzed on surface-based atlases. Cereb Cortex 22:2227-2240.
Van Essen DC, Donahue CJ, Coalson TS, Kennedy H, Hayashi T, Glasser MF (2019) Cerebral cortical folding, parcellation, and connectivity in humans, nonhuman primates, and mice. Proc Natl Acad Sci U S A.
Vazquez-Rodriguez B, Suarez LE, Markello RD, Shafiei G, Paquola C, Hagmann P, van den Heuvel MP, Bernhardt BC, Spreng RN, Misic B (2019) Gradients of structure-function tethering across neocortex. Proc Natl Acad Sci U S A 116:21219-21227.
Vogt C, Vogt O (1919) Allgemeinere ergebnisse unserer hirnforschung. J Psychol Neurol 25:279-462.
Voorhies W, Miller JA, Yao J, Bunge SA, Weiner KS (2020) Cognitive insights from evolutionarily new brain structures in prefrontal cortex. bioRxiv.
Voytek B, Kayser AS, Badre D, Fegen D, Chang EF, Crone NE, Parvizi J, Knight RT, D'Esposito M (2015) Oscillatory dynamics coordinating human frontal networks in support of goal maintenance. Nat Neurosci 18:1318-1324.
Waehnert MD, Dinse J, Weiss M, Streicher MN, Waehnert P, Geyer S, Turner R, Bazin PL (2014) Anatomically motivated modeling of cortical laminae. Neuroimage 93 Pt 2:210-220.
Wang L, Mruczek RE, Arcaro MJ, Kastner S (2015) Probabilistic Maps of Visual Topography in Human Cortex. Cereb Cortex 25:3911-3931.

Weiner KS, Natu VS, Grill-Spector K (2018) On object selectivity and the anatomy of the human fusiform gyrus. Neuroimage 173:604-609.

Weiner KS, Golarai G, Caspers J, Chuapoco MR, Mohlberg H, Zilles K, Amunts K, Grill-Spector K (2014) The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. Neuroimage 84:453-465.

Weiner KS, Barnett MA, Lorenz S, Caspers J, Stigliani A, Amunts K, Zilles K, Fischl B, Grill-Spector K (2017) The Cytoarchitecture of Domain-specific Regions in Human High-level Visual Cortex. Cereb Cortex 27:146-161.

Welker W (1990) Why does cerebral cortex fissure and fold? A review of determinants of gyri and sulci. In: (Peters A, Jones EG, eds), pp 3-136. New York: Plenum Press.

Yeo BT, Krienen FM, Eickhoff SB, Yaakub SN, Fox PT, Buckner RL, Asplund CL, Chee MW (2015) Functional Specialization and Flexibility in Human Association Cortex. Cereb Cortex 25:3654-3672.

Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zollei L, Polimeni JR, Fischl B, Liu H, Buckner RL (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 106:1125-1165.

Zilles K, Palomero-Gallagher N, Amunts K (2013) Development of cortical folding during evolution and ontogeny. Trends Neurosci 36:275-284.